

(12) **UK Patent Application** (19) **GB** (11) **2 216 530** (13) **A**
(43) Date of A publication 11.10.1989

(21) Application No 8906210.3

(22) Date of filing 17.03.1989

(30) Priority data

(31) 8806804

(32) 22.03.1988

(33) GB

(71) Applicant

**The Minister for Agriculture Fisheries and Food
Great Westminster House, Horseferry Road, London,
SW1P 2AE, United Kingdom**

(72) Inventors

**Bruce MacKenzie Pearson
Raymond Alexander McKee**

(74) Agent and/or Address for Service

**R F Walker
Ministry of Defence, Pats 1A4, Room 2014,
Empress State Building, Lillie Road, London,
SW6 1TR, United Kingdom**

(51) INT CL⁴

**C12N 15/00, C07D 495/04, C12N 1/18, C12P 11/00
// (C12N 1/18 C12R 1:865)**

(52) UK CL (Edition J)

**C3H HB7T HB7V H653 H654 H655 H685
C6Y Y125 Y330 Y501 Y503
U1S S1528**

(56) Documents cited

WO 87/01391 A1 JP 62155081 A JP 61202686 A

(58) Field of search

**UK CL (Edition J) C3H HB7E HB7T HB7V HB7X
INT CL⁴ C12N 15/00
ON-LINE - Biosis, CAB, Chemical Abstracts,
Derwent WPI.**

(54) Genetic material for expression of biotin synthetase enzymes

(57) Plasmids are provided which are capable of replication and expression in an organism other than *E. coli* and containing one or more genes derived from an *E. coli* *bioA*, *bioB* *bioC* *bioD* or *bioF* gene. Plasmids are also provided containing one or more of these derived genes in a form enabling their manipulation. Methods of preparing these plasmids are also described.

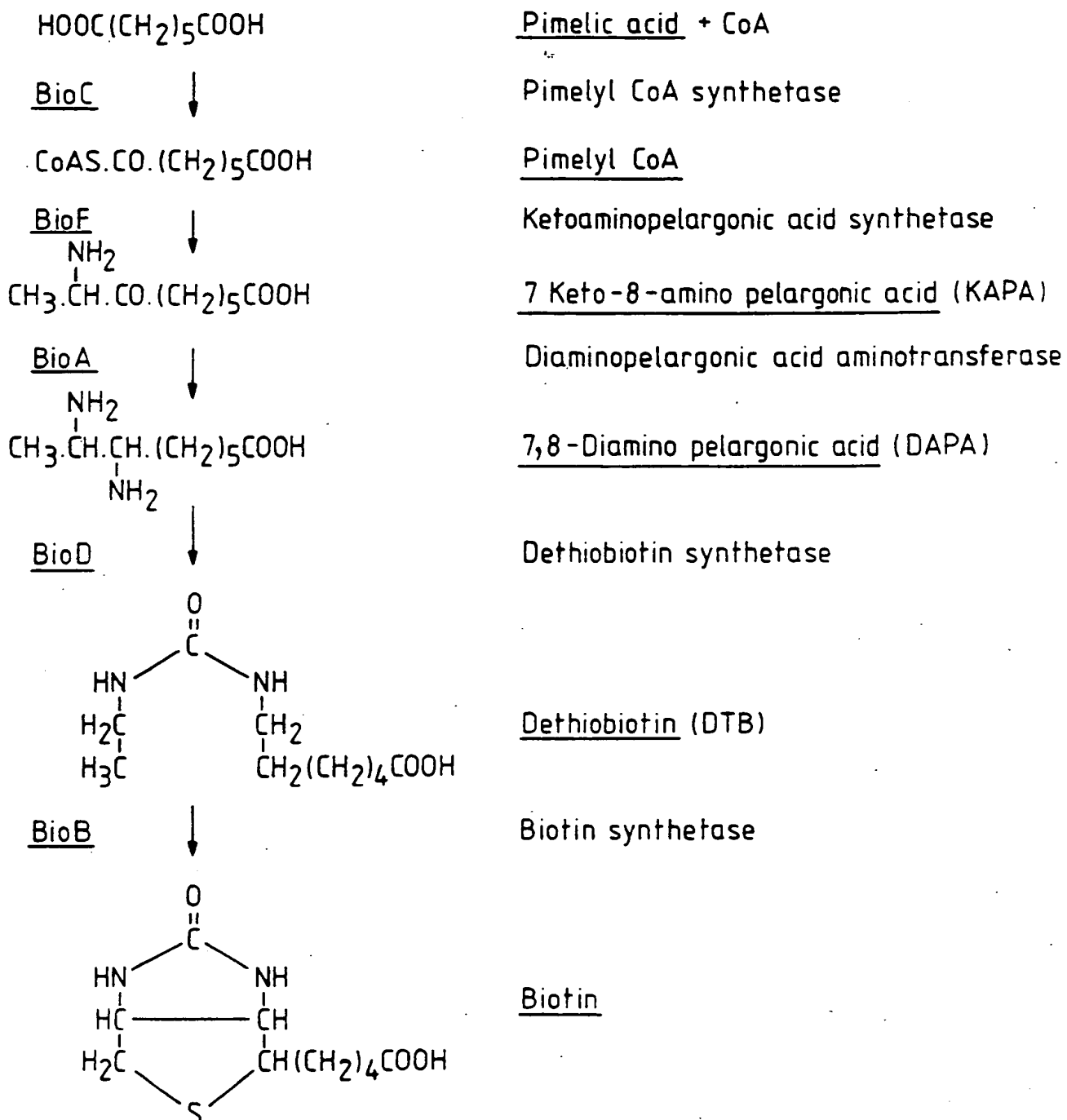
Fig. 1.

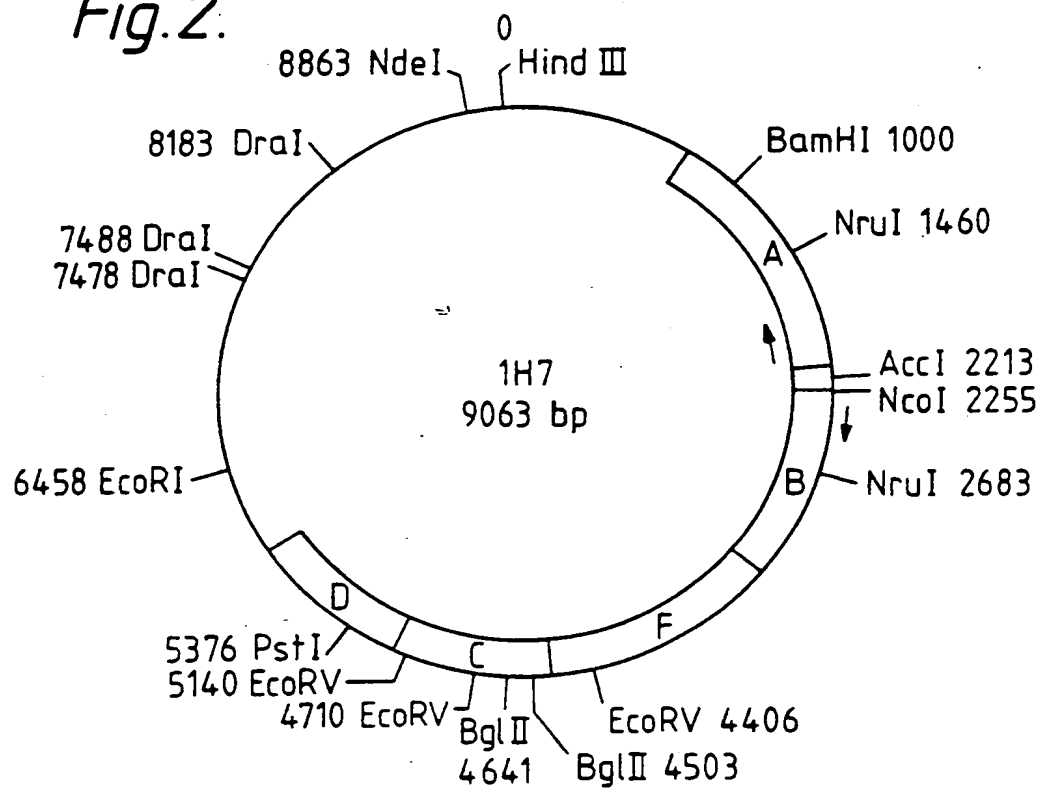
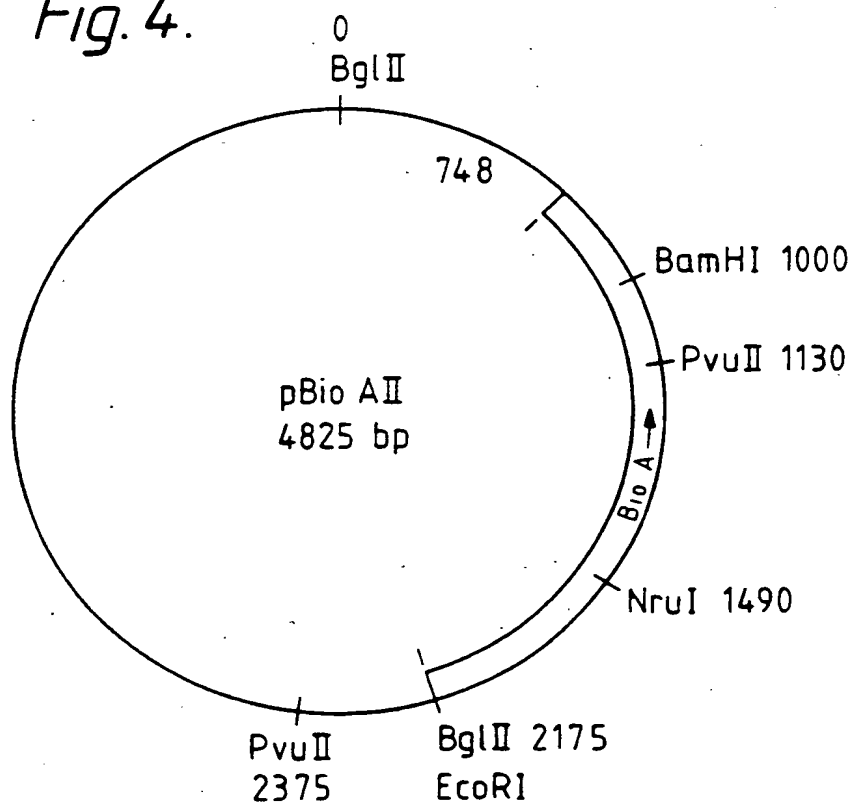
Fig. 2.*Fig. 4.*

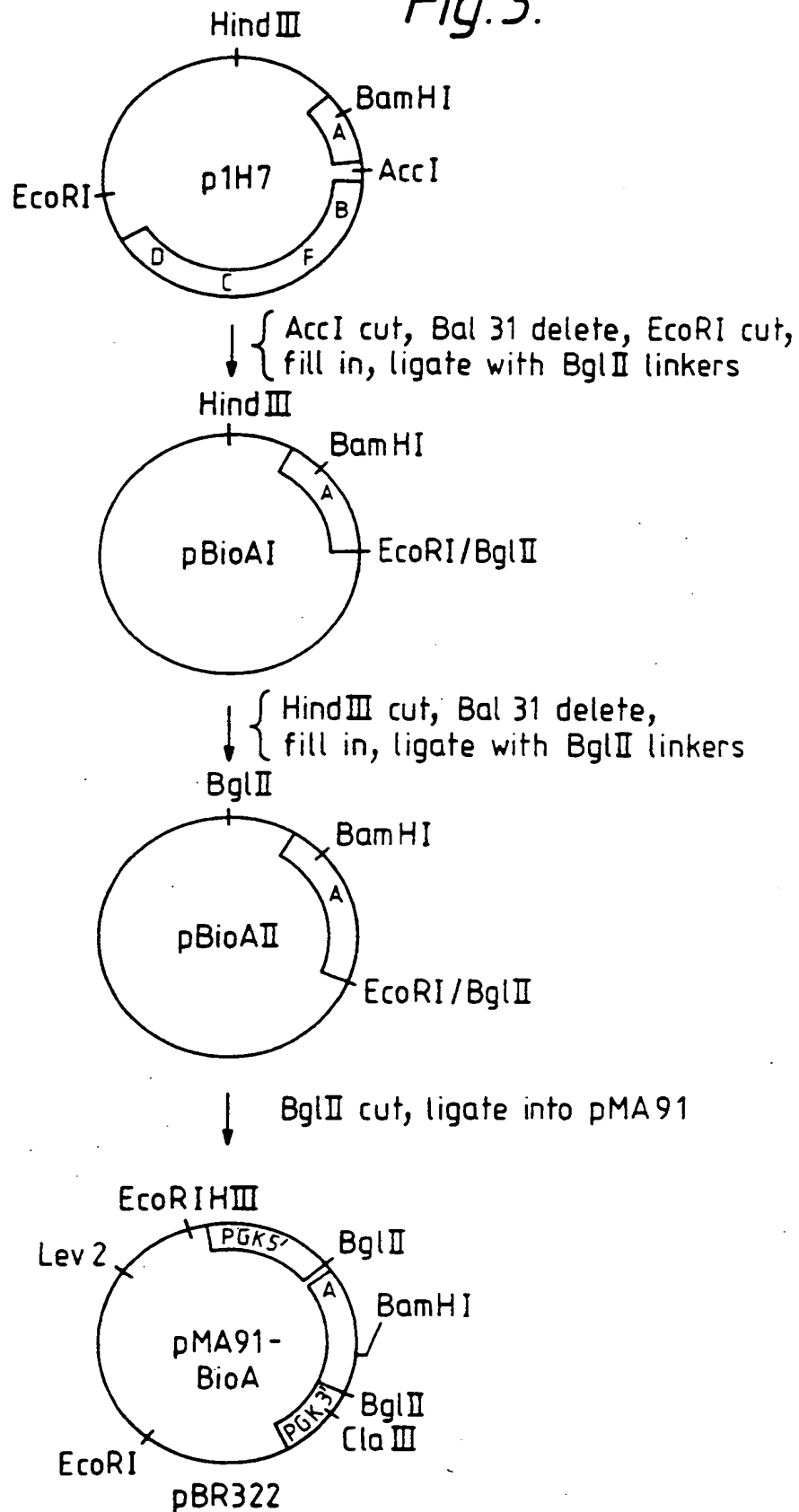
Fig. 3.

Fig. 5.

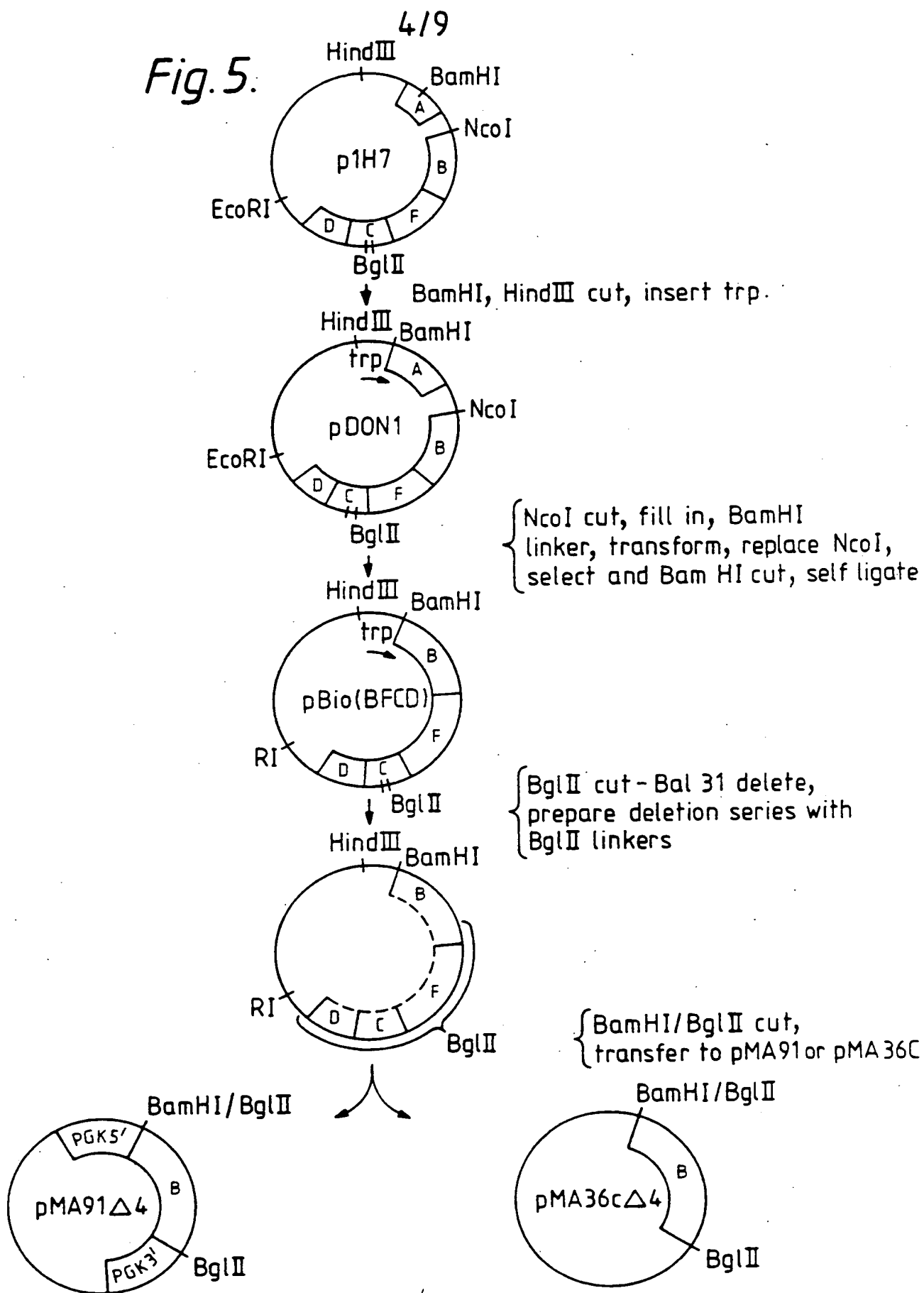


Fig.6.

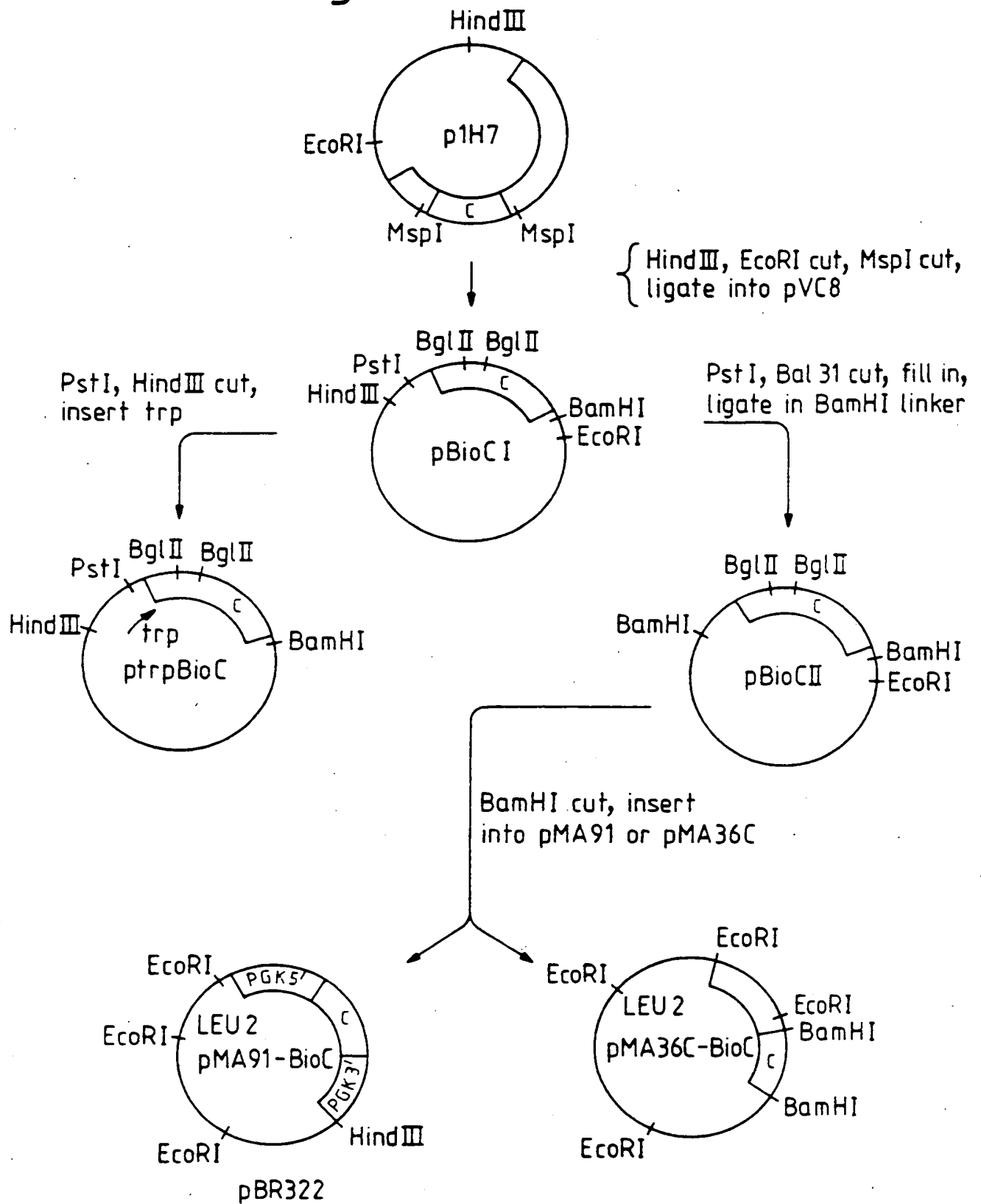
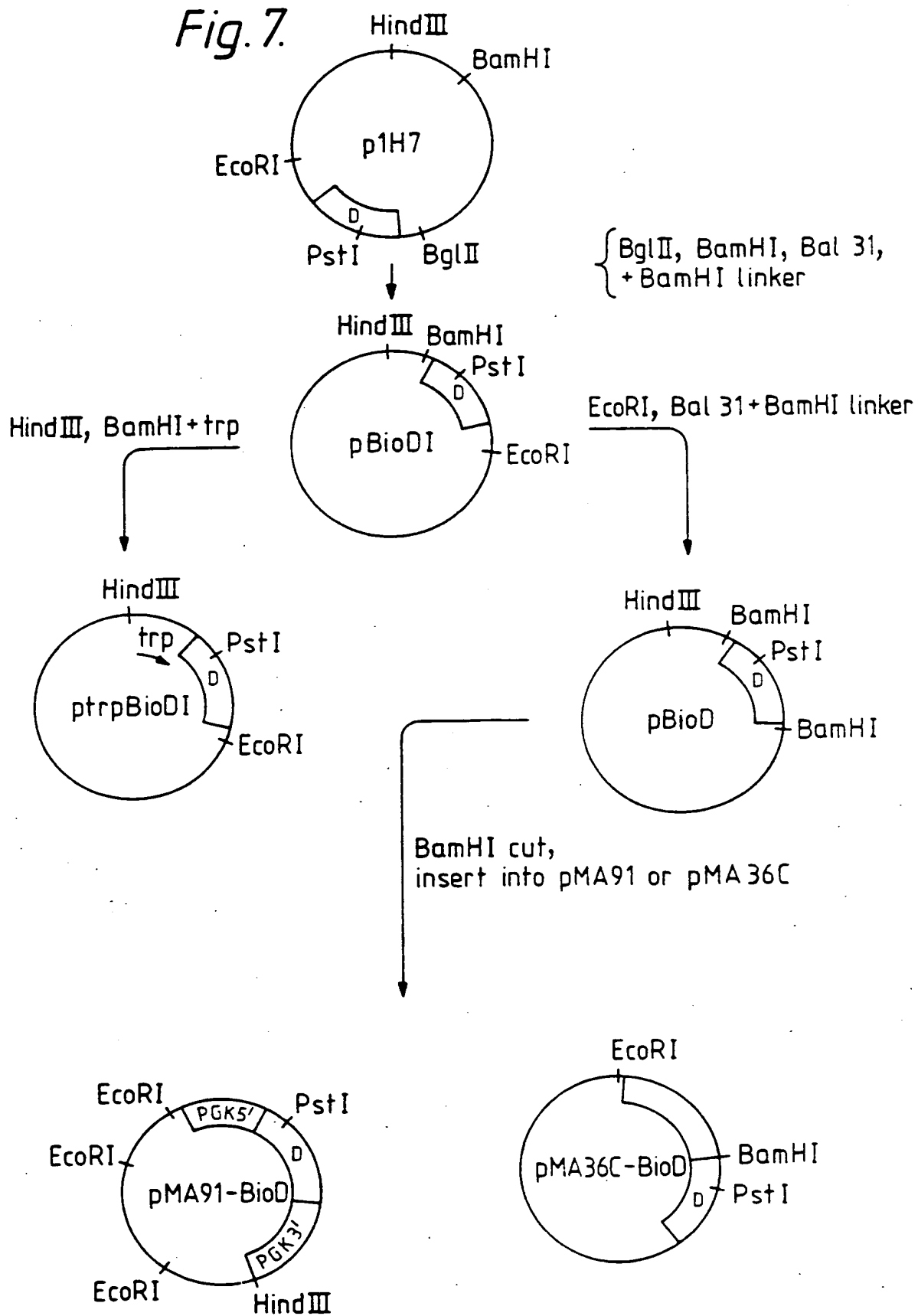


Fig. 7.



7/9

Fig. 8.

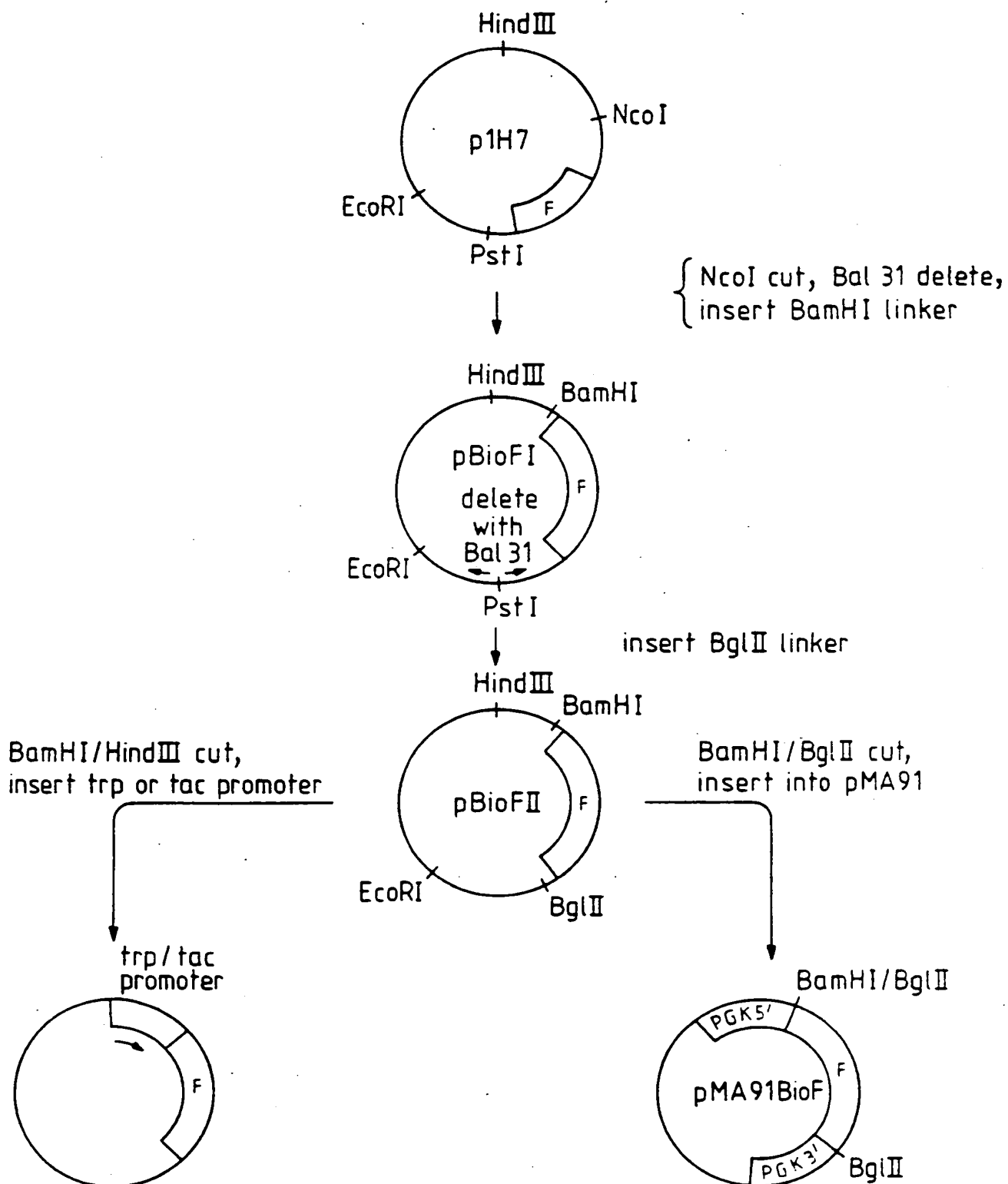


Fig. 9.

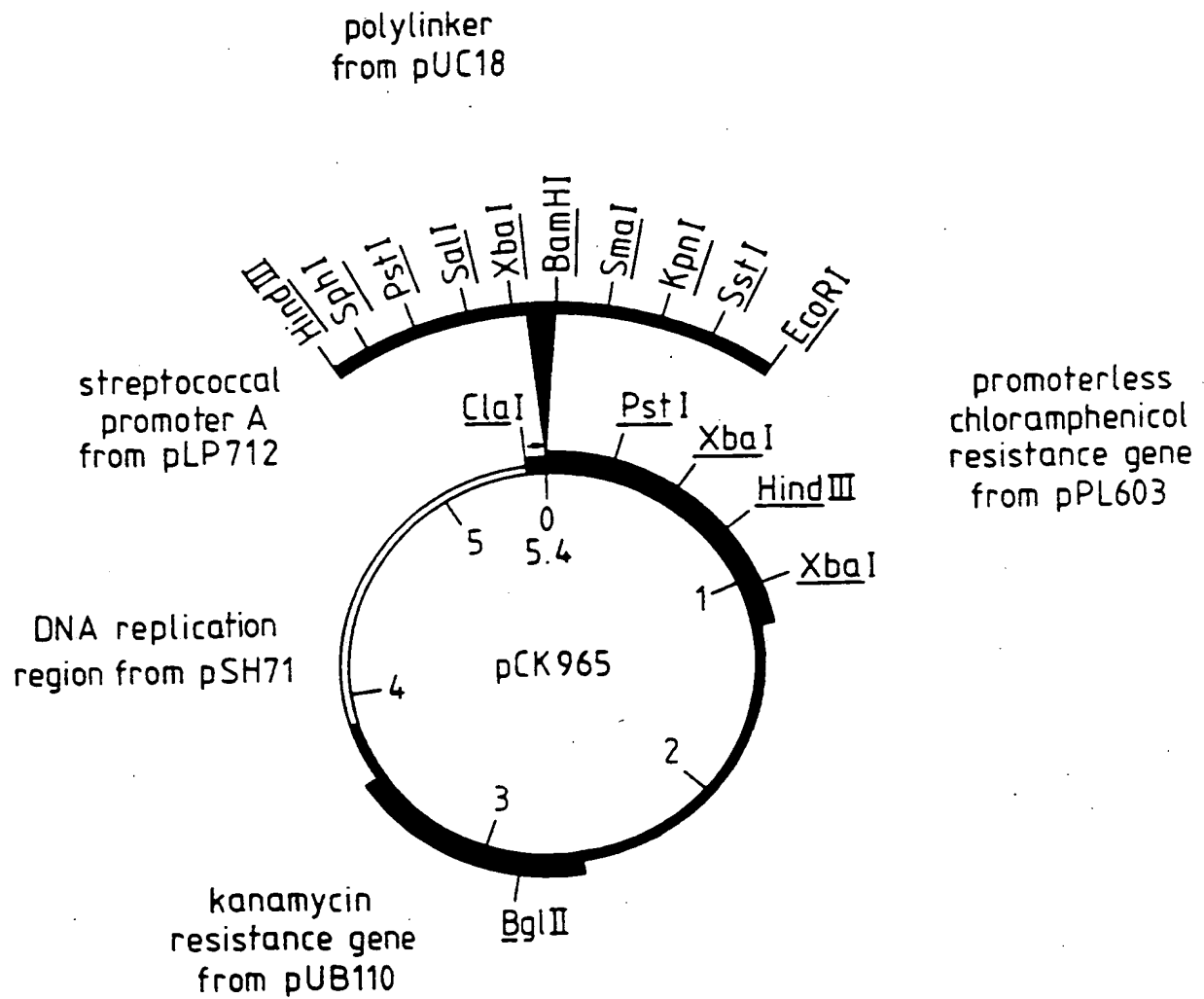
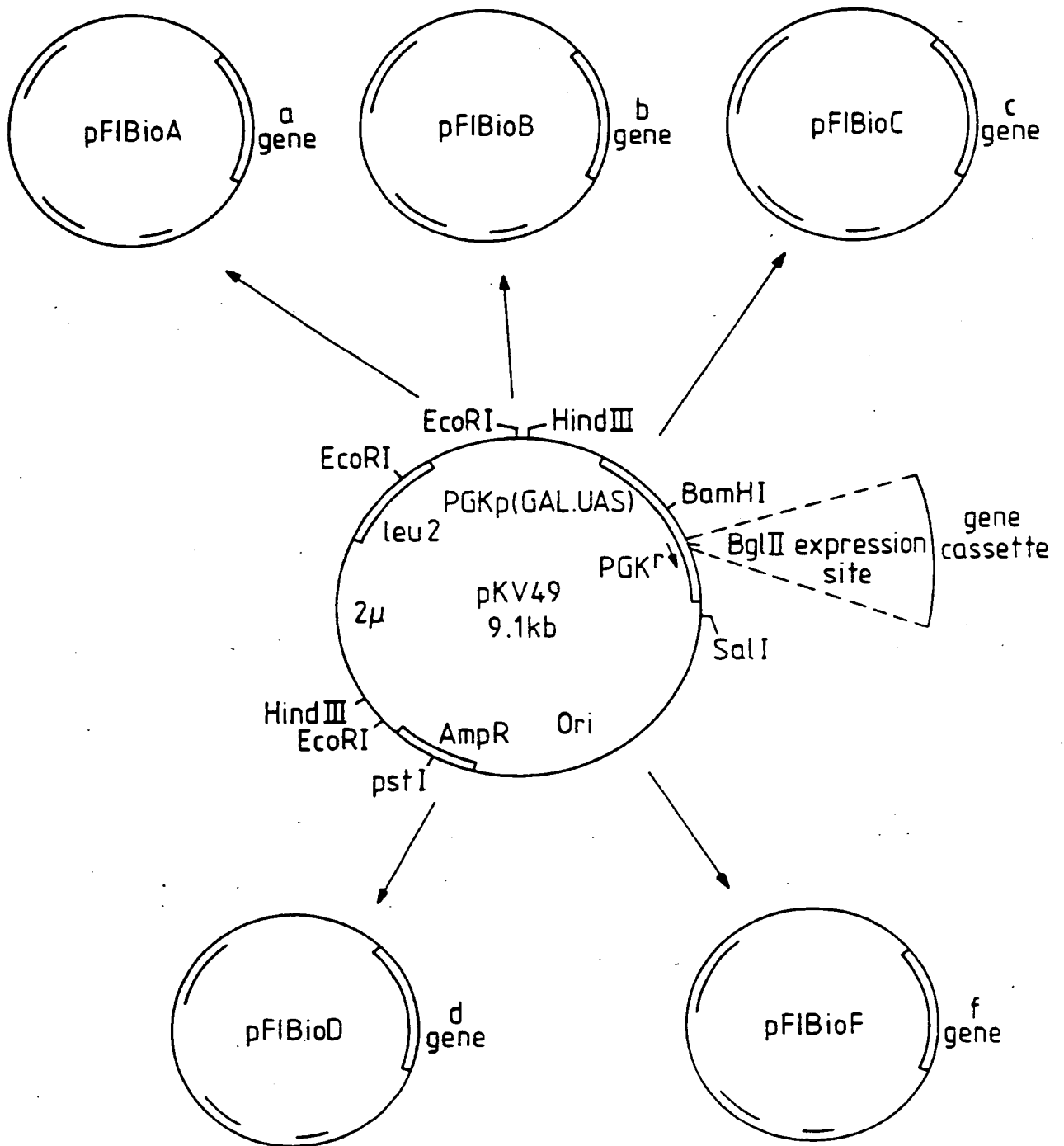
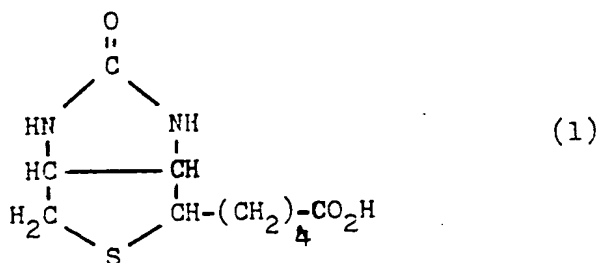


Fig. 10.

GENETIC MATERIAL FOR EXPRESSION OF
BIOTIN SYNTHETASE ENZYMES.

This invention relates to the use of recombinant genetic material (ie DNA sequences) for the expression of enzymes of the biotin synthetic pathway from micro-organisms, in particular from micro-organisms such as Escherichia coli. The invention also relates to micro-organisms which have been genetically transformed by introduction into them of such genetic material.

Biotin, (also known as vitamin H) has the structure (1):



It is probably an essential component of all living cells. Consequently it is a useful component of nutrients, especially of nutrients for those organisms which suffer from a deficiency of biotin, or which are unable to synthesise biotin effectively from their foodstuffs, for example certain yeasts and all mammals.

Biotin may be produced via synthetic organic chemistry but this is difficult and expensive. It is desirable to produce biotin biosynthetically from micro-organisms which are able to synthesise it from cheap starting materials such as glucose etc. In recent years the new technology of genetic engineering has become available whereby the genes of convenient micro-organisms may be modified by the introduction of genetic material which causes them to express a desired biosynthetic product which may then be harvested from a culture of the micro-organism.

Such genetic material is generally introduced into the micro-organism in the form of a plasmid, ie a circular length of extra-chromosomal DNA that is capable of autonomously replicating itself within the organism as the organism reproduces itself by cell division. Alternatively the genetic material may be integrated into the chromosome by recombination.

The prokaryotic micro-organism Escherichia coli is able to synthesise biotin from the simple precursor pimelic acid via the metabolic pathway shown in figure 1. In the course of this pathway five enzymes are involved, one for each biosynthetic step. These five enzymes are each the expression product of a portion of the biotin operon, the portions being described as the genes bioA, bioB, bioF, bioC, and bioD. The identity of the enzymes expressed by these genes is also shown in fig. 1. Herein "biotin synthetase" is used collectively for the expression products of all five genes of the biotin operon, but normally this term is used only for the product of bioB.

The entire biotin operon of E. coli, located between the bacteriophage λ attachment site and uvr-B gene loci at approximately 17 minutes on the E. coli genetic map (Bachmann and Low, 1980) has already been cloned from biotin-transducing λ phage (Das Gupta et al., 1978; Cohen et al., 1978). It comprises five closely linked complementation groups corresponding to the genes bioA, bioB, bioF, bioC and bioD which are divergently transcribed from two overlapping promoters (Cleary et al., 1972). This promoter region has been sequenced (Otsuka and Abelson, 1978; Barker et al., 1981). The leftwards transcript contains only the bioA gene whilst the rightwards transcript contains the remaining four genes (Guha et al., 1971) with bioB as the first translational product of the rightward transcript.

The bioB gene product has an estimated molecular weight of 36,000 Da (Dottin et al., 1975) and performs the dual function of catalyzing the incorporation of a sulphur atom into dethiobiotin with the loss of two protons and concomitant ring closure to form biotin.

Attempts have been made to improve the ability of E. coli to produce biotin by genetic modification. For example WO 87/01391 describes a method whereby the entire bioA - bioD sequence is isolated from an E. coli plasmid pLC2523, ligated with other DNA

and then reintroduced into E. coli. JP-A 60/42928 describes a similar method. JP-A 272605/84 describes a method whereby pLC 25-27 is cleaved to isolate a fragment containing an E. coli bioB and part of a bioF gene positioned between linkers, which is then ligated into various plasmids which are suitable vectors for transformation of E. coli. Evidence is presented in JP-A 272605/84 that the ability of E. coli to convert dethiobiotin to biotin is improved when that organism is transformed with these vectors.

Another class of micro-organism which is widely employed in industrial biotechnology, including the long established arts of brewing, winemaking, baking etc is the yeasts, such as Saccharomyces cerevisiae. Yeasts are eukaryotic organisms.

Sacch. cerevisiae is unable to synthesise biotin from simple precursors such as glucose or pimelic acid and therefore is not believed to contain all the enzymes of the normal biotin biosynthetic pathway. It does however, by inference from supplementation experiments with intermediates of the pathway in the absence of biotin, contain the final three enzymes, which perform a function equivalent to bioA, bioD and bioB of E. coli (Eisenberg 1973).

Early work on the conversion of dethiobiotin to biotin by this yeast (Niimura et al., 1964) demonstrated that exposure in the culture medium to dethiobiotin at concentrations greater than $2 \times 10^{-7} \text{M}$ ($\approx 50 \mu\text{g L}^{-1}$) decreased any subsequent biotin formation, a phenomenon which was ascribed to enzyme repression (Eisenberg, 1973). Although no more is known specifically about the regulation of biotin biosynthesis in Sacch. cerevisiae, it is clear that this step forms a major physiological block.

Eukaryotic organisms are substantially more complex than prokaryotic organisms, and this is reflected in the relative complexity, size and organisation of their genetic material. Whilst prokaryotes generally contain a single circular chromosome, eukaryotes often contain several chromosomes in their nucleus as well as mitochondrial DNA. There is therefore no prima facie reason to expect that a gene could be transplanted from a prokaryote into a eukaryote to achieve the same expressive effect in the latter.

The E. coli biotin operon is polycistronic. This means that the four genes of the rightwards transcript are translated in tandem form from the same RNA messenger molecule, rather than being produced individually. In addition there are likely to be attenuation sequences located between the genes of the E. coli operon which modify the quantitative efficiency of expression of each gene product. In this way the operon produces each biotin synthetase enzyme in a controlled quantity appropriate to the rate at which it performs its function so that biotin is synthesised at the rate at which E. coli requires it. By isolating the genes separately in the form of "cassettes" this level of control is removed and higher expression can be achieved.

Transcription in E. coli is controlled by the bir A gene of the E. coli chromosome. Additionally there are located in upstream positions relative to each E. coli biotin gene the appropriate shine delgano (SD) sequences which help to initiate the translational process carried out by the ribosome. The necessity for these DNA sequences in the E. coli biotin operon to control the activity of the five biotin genes suggests that these genes are precisely adapted to the synthesis of biotin by E. coli and that incorporation into other organisms would not produce a useful result.

Research into the effect of incorporation of the individual E. coli bio genes into organisms other than E. coli, and exploitation of such incorporation in industrial biosynthesis is at present hindered in practice by the unavailability of these individual genes in a form isolated from the other genes and from genetic material which is specific to E. coli.

It is an object of the invention to provide genetic material which may be introduced into microorganisms such as Sacch. cerevisiae to improve their ability to synthesise biotin. It is also an object of the invention to provide biotin genes in a form free of extraneous E. coli genetic material.

According to a first aspect of the invention there is provided a novel plasmid, capable of replication and expression in an organism other than E.coli, and containing one or more derived genes, being derived from an E. coli bioA, bioB, bioC, bioD or bioF gene.

5 By the term "derived gene" herein is meant a gene that expresses the same product or a product having the same or substantially the same biochemical function as that expressed by the gene from which it is derived, and which either has the same DNA sequence as the gene from which it is derived, or which has a high
10 degree of conformity to the DNA sequence of the gene from which it is derived, eg 70% or more, or which contains degenerate or preferred codons in place of the equivalent codons of the gene from which it is derived, or contains extra bases or lacks some bases relative to the gene from which it is derived.

15 In a preferred embodiment the term "derived from" includes a gene which is derived from the E. coli gene and has its codon sequence modified in favour of the other organism. Such modification is a consequence of the degeneracy of the genetic code, in that a number of codons may express the same amino acid, and whereas E. coli
20 may use preferred codons for each amino acid encoded in its genes, another organism may use a different but degenerate set of codons to express the same amino acids. A derived gene modified in favour of a particular organism therefore has such alternately preferred codons in place of those used by the gene from which it is derived.

25 An example of an E. coli bioB gene modified in favour of expression in the yeast Sacch. cerevisiae is listed in table 1.

The plasmid is preferably capable of replication and expression in yeasts, fungi, lactobacillus and other bacteria,
but especially in yeasts, particularly Sacch. cerevisiae and will
30 therefore contain suitable control DNA sequences operatively linked to the gene, including transcription / translation, promoter and terminator DNA sequences appropriate to the organism concerned. Suitable additional genetic material which may be combined with the gene to form a plasmid capable of replication and expression in a particular
35 micro-organism will depend upon the nature of the micro-organism. Certain plasmids of this aspect of the invention may also be capable of replication and expression in E. coli as well as the other organism.

In the case of yeasts, preferred additional genetic material is the known plasmids pMA91, pMA36c or preferably pKV49, and the plasmid may therefore consist of pMA91, pMA36c or pKV49 having the gene inserted into a suitable restriction site therein. In the case of *Lactobacillus*, suitable genetic material includes the known plasmid pCK965, and the plasmid may therefore consist of this plasmid having the gene inserted into a suitable restriction site, eg the polylinker site therein. The restriction map of pCK965 is shown in Fig 9.

Other known plasmids which are capable of replication in yeast *lactobacillus* or other organisms may also be used. Suitable transcription/translation promoters for use in yeast include Trp, Gap, Pho 5, Gal 1-10, Pal and PGK.

In the plasmid of the first aspect of the invention, the gene(s) may be linked at the end(s) into the plasmid via known linkers or adaptors. Conveniently these may be linkers or adaptors which allow easy excision of the gene from the plasmid by the use of appropriate restriction enzymes, and reinsertion of the gene into other genetic material, eg. other plasmids, by the use of suitable ligating enzymes. In this embodiment of the invention the plasmid provides the gene in a form which allows easy insertion into other micro-organisms, via appropriate plasmids. Particularly suitable linkers are BglIII and/or BamHI linkers.

As well as containing the gene, linkers, expression and replication operators, the plasmid may also contain other genetic material which is coding or non-coding, for example genetic markers to enable easy selection during cloning operations eg. resistance to antibiotics such as ampicillin or tetracycline.

In a preferred embodiment of this first aspect of the invention the plasmid is substantially free of any *E. coli* genetic material from the biotin operon which is other than directly coding for the bioA, bioB, bioC, bioD or bioF gene or genes contained in the plasmid. It is particularly preferred in this case that the plasmid contains no *E. coli* promoters, attenuation sequences, Shine Dalgarno sequences or fragments acting on these *E. coli* genes. When the plasmid contains only one bioA, B, C, D or F gene it is particularly desirable that no fragments of the other genes are present in the plasmid, and further that there should be no ATG base sequence between the linker or adaptor 5' to the start codon of the gene.

The novel plasmids of the first aspect of the invention may be used as vectors, that is they may be introduced into a micro-organism in which they will replicate and express their gene product, in this case one of the biotin synthetase enzymes.

5 Introduction of such a vector into a micro-organism may usefully modify the biosynthetic pathway by which biotin or its precursors ("vitamers") are synthesised within the organism. For example introduction of vectors containing bioC and bioF genes into yeast can be used to enable the yeast to grow in the absence of
 10 biotin, for example by enabling it to synthesise the two precursors pimelyl CoA and 7-keto-8-aminopelargonic acid. This has the potential for use as a novel selective marker for yeasts in fermentation. Many known selective markers rely on the presence of resistance to certain antibiotics in the strain being selected. Insertion of bioC
 15 and bioF genes would enable the useful possibility of selection of a strain that will grow in the absence of biotin. Similarly insertion of bioA, bioB and/or bioD genes into yeast can supplement the activity of the corresponding genes already present in the yeast. In such ways the yield of biotin from an organism which
 20 exports biotin may be improved, or intracellular levels increased, in some cases substantially, and the ability of an organism which lacks the ability to synthesise biotin for its own use to grow in a medium containing little or no biotin may be improved.

25 In the course of synthesis of the plasmid of the first aspect of the invention (see below) other useful novel plasmids are provided. In particular according to a second aspect of the invention there is provided a plasmid containing one or more derived genes, being derived from an E. coli bioA bioB bioC bioD or bioF gene and containing no E. coli control sequences in a position to act on
 30 said derived gene. The derived gene is preferably linked into the plasmid via synthetic linkers.

5 As well as containing the gene and the linkers, the plasmid of
 this second aspect of the invention may also contain additional DNA, eg
 it may contain control sequences located upstream and/or downstream
 of the gene such as the yeast promoters mentioned above. Preferably
 the gene is linked into the plasmid by linkers or adaptors which are
 10 positioned as close as possible to the 3' and 5' end of the gene, so
 that as little non-coding DNA as possible is included with the gene
 and any control sequences which are present between the linkers.
 Preferably linkers are chosen that may be conveniently cleaved by
 restriction endonuclease enzymes, to excise the gene embodied in a
 15 DNA sequence having ends which may then be conveniently ligated into
 other genetic material, such as the yeast pMA91, pMA36c and preferred
 pKV49 plasmids referred to above, or into E.coli expression
 vectors with promoters such as those of pKK223-3 or pDR720 available
 from Pharmacia LKB Biotechnology, S-751 82 Uppsala, Sweden. Preferred
 20 linkers are Bam HI and Bgl II.

The plasmid of this second aspect of the invention advantageously
 provides the gene it contains in the form of a "cassette" which can
 be easily excised and then reinserted into other genetic material.
 If the regulatory control signals are absent from the 3' and 5' ends
 25 the gene may be inserted in different transcriptional orientations
 relative to their orientation in the E. coli chromosome, and
 combined with other regulatory signals to produce an artificial
 operon.

The plasmid of the first or second aspects of this invention
 30 may contain one or more of the derived genes referred to above, but
 conveniently to enable the properties of a single gene to be
 investigated it contains only one such gene, for example bioB, bioC,
bioF etc.

Using such a cassette, the gene(s) it contains may be inserted
 35 into for example other plasmids (which may be plasmids of the first
 aspect of the invention) or into the chromosome of an organism.

The plasmids of the above aspects of the invention may be used to transform microorganisms, using known transformation techniques. As a further aspect of the invention therefore a novel microorganism is provided, comprising a known microorganism which has been transformed by insertion of a plasmid of the invention. The known microorganism may for example be a yeast, lactobacillus or other bacillus or E. coli, but is preferably a yeast, for example Sacch. cerevisiae, Yarrowia lipolytica and Pichia sp.

10 Plasmids according to the first and second aspects of the invention may be prepared by the method generally described below:

(1) The known phage Charon 4A is used to take up the whole of the biotin operon, containing the bioA, bioB, bioF, bioC and bioD genes from E. coli.

15 (2) The DNA containing the biotin operon is isolated from the phage, for example by phenol extraction and ethanol precipitation.

(3) A length of the DNA known to contain the operon is excised using restriction enzymes, for example EcoRI and Hind III, and the fragment containing the operon is isolated, for example by gel electrophoresis.

20 (4) The fragment from (3) is combined with a suitable plasmid for cloning eg pUC8. The combined plasmid and fragment are cloned. Clones containing the combination are selected, and the combined DNA is isolated.

25 (5) By the use of appropriate known restriction endonuclease enzymes, ligation enzymes and cloning techniques, including sequencing and the polymerase chain reaction ("PCR"), the five genes bioA, bioB, bioC, bioD and bioF may be separated from each other and from E. coli DNA and inserted into other genetic material to form the novel plasmids of the invention.

30 A gene having its sequence modified in favour of another organism may be synthesised using known DNA synthesis techniques. It may then have suitable synthetic linkers such as Bam HI and/or Bgl II ligated on at its 3' and 5' ends, by means of which it may be inserted into genetic material to form a plasmid of the first and second aspects of the invention using conventional techniques.

The invention will now be described by way of example only with reference to the following figures.

- Fig 1. shows the biosynthetic pathway for synthesis of biotin from pimelic acid. The role of the biotin synthetase genes is shown together with the expression product of the genes.
- Fig 2. shows the restriction map of plasmid pLH7.
- Fig 3. shows the procedure for preparation of plasmids pBioAI, pBioAII and a plasmid pMA91-BioA.
- Fig 4. shows the restriction map of plasmid pBioAII.
- Fig 5. shows the preparation of plasmids pMA91 Δ 4 and pMA36 Δ 4.
- Fig 6. shows the preparation of plasmids ptrpBioC, pMA91BioC and pMA36BioC.
- Fig 7. shows the preparation of plasmids ptrpBioDI, pMA91BioD and pMA36BioD.
- Fig 8. shows the preparation of plasmids pMA91BioF and pMA36BioF.
- Fig 9. shows the restriction map of plasmid pCK965, which is a lactobacillus expression vector.
- Fig 10. shows the incorporation of the individual bio A, bio B, bio C, bio D, bio F genes into a yeast expression vector pKV 49.

In the following examples the following materials and methods were employed.

Microorganism strains

- Escherichia coli strain JM 83 (known strain: ACTT 35607).
(ara, Δ (lac-proAB) thi, strA, \emptyset 80, lacZm M15) used routinely was kindly provided by R James of the University of East Anglia, Norwich, UK, whilst Saccharomyces cerevisiae (NCYC 1527 (α) (leu2-3, leu2-112, ura 3, his3-11, his3-15 trp1) was obtained from the National Collection of Yeast Cultures, AFRC Institute of Food Research, Norwich.

Culture media

Escherichia coli was grown routinely in LB medium or a minimal medium with or without additions of ampicillin ($100\mu\text{g ml}^{-1}$), β -indoleacrylic acid ($5\mu\text{g ml}^{-1}$), dethiobiotin ($2\mu\text{g ml}^{-1}$) or avidin ($0.1\mu\text{g ml}^{-1}$) as required. Saccharomyces cerevisiae was grown either on YEPD or Vit-2 medium. The composition of the media were as follows (per L of water). LB: Bacto tryptone (10 g),

Bacto yeast extract (5g), NaCl (10g), pH 7.5. Minimal medium: glucose (2g), vitamin assay casamino acids (2g), Na_2HPO_4 (7g), KH_2PO_4 (3g), NaCl (0.5g), NH_4Cl (1g) Na_2SO_4 (0.8g), MgSO_4 (0.3g), CaCl_2 (20mg), thiamine HCl (20mg) histidine HCl (20mg), thymine (40mg), proline (20mg), diaminopimelic acid (100mg), thymidine (40mg). YEPD: Bacto yeast extract (10g), Bactopeptone (20g), glucose (20g). Vit-2 medium: Bacto vitamin-free yeast base (16.7g), inositol (10mg), Ca pantothenate (2mg), pyridoxine HCl (0.4mg) thiamine HCl (0.4mg), nicotinic acid (0.4mg), para-aminobenzoic acid (0.2mg), riboflavin (0.2 mg), folic acid (2 μg).

Plasmids and phage

Plasmid pLH7 containing the whole biotin operon was prepared in our laboratory by insertion of an EcoRI-Hind III fragment from the λ -bio transducing phage Charon 4A (obtained from N Ellis AFRC, John Innes Institute, Norwich, UK) into the complementary polylinker sites of pUC8 (Vieria and Messing, 1982). pMA91 was obtained from A J and S Kingsman, Dept Biochemistry, University of Oxford, Oxford, UK.

DNA sequence determination

Sequencing was performed by the Sanger dideoxy method (Sanger et al., 1977; 1980) using ^{35}S -labelled nucleotides. Briefly, random restriction fragments obtained by digestion with Sau3A, Taq I and Hpa II of the EcoRI/HindIII fragment of plasmid LH7 containing the entire operon isolated from an agarose gel by freeze squeezing (Tautz & Renz, 1983) were cloned into the appropriate polylinker sites of M13mpl8 and/or M13mpl9. Similarly, the C-terminus of the bioB gene was confirmed by sequencing from the Bgl II linker inserted into a deletion series prepared by Bal 31 digestion of Bgl II cut LH7.

Sequence data analysis was done by the DNA-sequencing program DNASTAR for IBM personal computers by DNASTAR Inc. 1801 University Avenue, Madison WI, 53705.

Measurement of biotin

- 5 Biotin was measured in the specific microbiological assay employing Lactobacillus plantarum NCIB 6376 by the method of Wright and Skeggs (1944).

Preparation of plasmids containing biotin genes

10 1. Preparation of plasmid pLH7

- The transducing phage charon^{4A} contains the whole biotin operon from E.coli. The phage is phenol extracted with phenol equilibrated with TE(10mM Tris-HCl, 1mM EDTA pH 8.0), precipitated with 0.1 volume 3M sodium acetate pH 7.4 and
 15 2 volumes of ethanol, left at -20° for 30 minutes and centrifuging at 12,000g for 30 minutes at 4°C. The pellet was washed with 70% ethanol, dried in a vacuum and resuspended in TE. The DNA (10µg) is then cut with 5 units of the restriction endonucleases EcoRI and HindIII for three hours
 20 at 37°C in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl. The fragments are run on a 1% agarose gel in TAE buffer, stained with ethidium bromide 1 µg/ml and visualised under UV. The band at approx 6.3kb (using lambda HindIII as standard) was cut out and eluted. The DNA is then ethanol precipitated and resuspended
 25 in TE.

- The plasmid pUC8 (10µg) is digested with EcoRI and HindIII as above, phenol extracted, ethanol precipitated and resuspended in TE. The digested pUC8 and the 6.3kb fragment are then ligated in ligation buffer with 1 unit of T₄ ligase
 30 at 14°C for 12 hours. This is then used to transform E.coli JM83, and ampicillin resistant white colonies (β -galactosidase negative, white on X-gal medium) grown up and screened for the presence of the 6.3kb fragment following digestion with HindIII

and EcoRI is called pIH7, and its structure is shown in Fig 2.

Plasmid pIH7 was used as the starting point for preparation of plasmids containing the other biotin synthetase genes.

5 2. Preparation of plasmids containing the BioA gene.

10 The plasmid pIH7 (10 µg) was cut with 5 units of the restriction endonuclease AccI at 37°C for 3 hours in 50mM Tris-HCl pH 8.0, 10 mM MgCl₂, 100 mM NaCl, phenol extracted and ethanol precipitated, washed with 70% ethanol and vacuum dried. The plasmid was resuspended and
15 treated with the exonuclease Bal31 in 20mM Tris-HCl pH 8.0, 12 mM MgCl₂, 12mM CaCl₂, 600 mM NaCl, 1mM EDTA at 15°C, until approx 100 bp were removed. The mix was then phenol extracted and ethanol precipitated. The plasmid was resuspended and treated with 5 units of restriction endonuclease EcoRI at 37°C for 3 hours in 50 mM Tris -
20 HCl pH 8.0, 10mM MgCl₂, 100mM NaCl, phenol extracted and ethanol precipitated.

The ends were filled with DNA polymerase polymerase klenow fragment and blunt end ligated with a BglII linker. This removes the other genes from the operon leaving BioA. The plasmid is now designated pBioAI and is cut with the restriction endonuclease HindIII, phenol extracted and ethanol precipitated. The plasmid is resuspended and treated with Bal31 exonuclease in 20 mM Tris-HCl pH 8.0, 12 mM MgCl₂, 12 mM CaCl₂, 600 mM NaCl, 1 mM EDTA at 15°C to remove non-bioA coding DNA, and the ends are filled in using DNA polymerase klenow fragment. Another BglII linker is ligated and the construct used to transform the E. coli JM83.

The ampicillin resistant colonies are tested for plasmid containing two BglII sites and now termed pBioAII. The BioA gene is now able to be inserted as a cassette. By ligating the BglII fragment containing the BioA gene into BglII digested pMA91 or BamHI digested pMA36C we have novel plasmids, for expression in yeast. This procedure is shown in Fig 3, a more complete restriction map of pBioAII being shown in Fig 4.

This BioA gene will supplement the gene already present in yeast to increase conversion of 7-keto-8-amino-pelargonic acid to 7,8 diaminopelargonic acid. The fragment is also suitable for expression in other yeasts, lactobacillus and bacillus species using appropriate vectors and promoters. The gene has been partially sequenced using the dideoxy chain termination method. The part sequenced contains the BamHI site (see table 2).

3. Preparation of plasmids containing the BioB gene

The plasmid pLH7 prepared in 1 above was digested with the restriction endonucleases BamHI and Hind III and the trp promoter cartridge is ligated in to form pDON1.

The BioB gene has a unique NcoI restriction site at the start. The pDON1 plasmid (10µg) was digested with five units of NcoI for three hours at 37°C in 50mM Tris-HCl pH8.0, 10mM MgCl₂, 50mM NaCl, phenol extracted and ethanol

precipitated. The pellet was washed with 70% ethanol, vacuum dried and resuspended in klenow polymerase buffer. The ends were filled in with DNA polymerase klenow fragment and ligated in the presence of an eight base pair BamHI linker. The colonies are checked for the presence of two BamHI sites and those digested with BamHI, self ligated and screened for the presence of a single BamHI site.

The plasmid pBioBFCD is now digested with BglII and incubated with the exonuclease Bal31 for 20 minutes. The single stranded ends are filled in using DNA polymerase klenow fragment and ligated in the presence of a BglII linker. Following transformation, colonies are screened for a 1070bp BamHI-BglII fragment indicating the isolation of the BioB gene. Selection was on the basis of sequence analysis, the DNA samples being subcloned into M13mpl9 virus and sequenced using the dideoxy chain termination method. Clones with the minimum 3' sequence were chosen for use. The sequence was found to be that of JPA 272605/48, ie as listed in table 3.

By ligating the BamHI-BglII fragment into BglII digested pMA91 or BamHI digested pMA36C novel plasmids pMA Δ 4 and pMA36C Δ 4 were prepared, suitable for expression in yeast. pMA91 expresses at a higher level than pMA36C. The insertion method is shown overall in Fig 5, and is described below.

Method

Clone Δ 4 from the pBioB-D series (produced by cutting pKH7 at NcoI, ligating in BamHI linker, BamHI digest, self ligate, then BamHI/HindIII and insert a trp promoter. Digest with BglII, treat with Bal31 and put in a BglII linker. The bioB gene was removed by cutting Δ 4 with BamHI/BglII, running the digest on a mini gel and removing the small fragment.

The strip of gel containing the bioB gene was set in the bottom of a 1% agarose gel and a piece of DE 81 filter paper inserted at the bottom of the strip and the DNA run into that overnight by electrophoresis. The DNA was then extracted from the paper in 1.5 M NaCl. 1mM EDTA, phenol extracted and ethanol precipitated and dissolved in 10ul water.

Then both the pMA91 BglIII and pMA36C BamHI digests were ligated separately overnight at 14°C with the BamI/BglIII Δ 4 digested DNA. The product was phenol extracted, ethanol precipitated, resuspended in 10 μ l water and 2 μ l used to transform competent JM83.

5 The ligation medium below was used:

1 μ l pMA91 (Bg lII) pMA36c (Bam H1)

2 μ l Δ 4 (Bam H1 digested)

2 μ l x 5 ligm. buffer

1 μ l ligase

10 4 μ l water

Mini preps of DNA were made from the transformants and the DNA was cut with BamHI/HindIII. pMA36C Δ 4 was seen to have an insert in the correct orientation.

DNA was prepared from the transformants and cut with HindIII/BamHI to see if an insert was present, and the orientation of these were investigated by cutting with HindIII and HindIII/BglII.

This BioB gene will supplement the gene already present to increase conversion of dethiobiotin into d-biotin by yeast in culture. The fragment is also suitable for expression in other yeasts, lactobacillus and bacillus using appropriate vectors incorporating control sequences for expression in these three organisms.

4. Preparation of plasmids containing the BioC gene

The gene is situated between BioF and BioD in the biotin of E.coli. Plasmid pLH7(10 μ g) was cut with five units of the restriction endonucleases HindIII and EcoRI in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl and incubated at 37°C. After three hours the DNA was run on a 1% agarose gel in TAE buffer, stained with ethidium bromide (1 μ g/ml), visualised under UV and the band at 6.3kb cut out. The DNA was eluted from the gel, ethanol precipitated and resuspended in TE. The DNA is then digested with the restriction endonuclease MspI in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 50mM NaCl and incubated at 37°C for three hours. The BioC gene is bounded by two Msp I sites and the fragment is about 900bp in length. The digested DNA is run on an 1.2% agarose gel in TAE buffer, stained with ethidium

bromide (1µg/ml) and visualised under UV, and the band at approx. 772 bp cut out and eluted. The fragment was then ethanol precipitated, vacuum dried and resuspended in TE.

The plasmid pUC8(10µg) was digested with five units of the
 5 restriction endonuclease AccI in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl and incubated at 37°C for three hours, phenol extracted, ethanol precipitated and ligated with the BioC containing fragment in a 1:2 ratio. E.coli JM83 was transformed and plasmid prepared from white ampicillin resistant
 10 colonies growing on LB amp plus X-Gal. These were checked for the presence and orientation of the BioC gene by cutting with restriction endonucleases HindIII and BglII. This gave plasmid pBio CI.

Plasmid pBioCI (10µg) was digested with 5 units of restriction
 15 endonucleases PstI and HindIII in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100 mM NaCl and incubated for 3 hours, phenol extracted, ethanol precipitated and resuspended in TE. The trp promoter was inserted giving plasmid ptrpBioC which is capable of expressing pimelyl CoA synthetase activity in E.coli.

20 Plasmid pBioCl was digested with the restriction endonuclease PstI in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100 mM NaCl and incubated for 3 hours, phenol extracted, ethanol precipitated and resuspended in TE followed by Bal31, to remove approximately 200 bp, filled in with DNA polymerase klenow
 25 fragment and ligated with a BamHI linker. BamHI digestion of this plasmid now provides a cassette for the movement of this gene into other organisms for expression. For example into BglII digested pMA91 or BglII digested pMA36C for expression in yeast. This procedure is shown in Fig 6. The sequence of this BioC gene is
 30 shown in table 4.

Other linkers may be put at either or both ends of the BioC gene to provide a cassette for insertion into plasmids suitable for transformation of other microorganisms.

5. Preparation of plasmids containing the BioD gene

35 The plasmid pKH7 (10µg) is cut with five units of each of the restriction endonucleases BglII and BamHI in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl and incubated at

37°C for three hours. The DNA is run on a 1% agarose gel stained with ethidium bromide (1µg/ml) visualised with UV and the DNA in the band at 5390bp removed.

5 The DNA is treated with the exonuclease Bal31 in 20mM Tris-HCl pH 8.0, 12mM MgCl₂, 12mM CaCl₂, 600mM NaCl, 1mM EDTA at 15°C, ends filled in with DNA polymerase klenow fragment and ligated in the presence of a BamHI linker. Following transformation plasmid is prepared from the ampicillin resistant colonies and checked by cutting with BamHI in 50mM
10 Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl and incubated at 37°C for three hours. Those with linker inserted then have their BamHI PstI fragment subcloned into M13mpl9 virus and sequenced using the dideoxy chain termination method. The start of the gene was identified and the clone with its Shine
15 Delgarno sequence still intact was designated plasmid pBioDI.

Part of the Bio D gene is listed in table 5.

The Trp and Tac promoter can be used for expression in E.coli by inserting these in the correct orientation in place of the 1kb BamHI HindIII fragment from pBioDI. The plasmid is
20 cut at the EcoRI site, phenol extracted, ethanol precipitated, treated with Bal31 exonuclease in 20mM Tris-HCl pH 8.0, 12 mM MgCl₂, 12mM CaCl₂, 600 mM NaCl, 1mM EDTA at 15°C to remove approximately 200 bp, and the ends are filled in using DNA polymerase klenow fragment. A BamHI linker is inserted and this provides a cassette for expression of dethiobiotin synthetase in
25 yeast, lactobacillus or bacillus under appropriate promoter control.

By ligating the BamHI-BamHI fragment into pMA91 or pMA36C as described above a novel plasmid for expression in yeast is provided. This procedure is shown in Fig 7.

30 6. Preparation of plasmids containing the BioF gene

The plasmid pLH7 (10µg) was digested with 5 units of the restriction endonuclease NcoI in 50mM Tris-HCl pH 8.0, 10 mM

MgCl₂, 50mM NaCl and incubated at 37°C for three hours. It was
 then phenol extracted, ethanol precipitated and resuspended in
 20mM Tris-HCl pH 8.0, 12mM MgCl₂, 12mM CaCl₂, 600mM NaCl,
 then reacted at 15°C with the exonuclease Bal31. After removal
 5 of approx 2Kb the reaction was stopped by the addition of EDTA,
 phenol extracted, ethanol precipitated, vacuum dried and
 resuspended in TE. The DNA was ligated in the presence of a
 BamHI linker and transformed into the E.coli JM83 and ampicillin
 resistant colonies selected. Transformants had their plasmid
 10 DNA (10µg) isolated and this cut with the restriction endo-
 nucleases BamHI and PstI to indicate the length of deletion
 and then appropriate clones sequenced by subcloning into
 BamHI- PstI cut M13mp19 virus and sequenced using the dideoxy
 chain termination method. A clone was found which had the
 15 Shine Delgarno sequence intact but none of the BioB gene
 remained 5' to this. The plasmid (10µg) was cut with
 5 units of the endonuclease PstI at 37°C for three hours
 in 50mM Tris-HCl pH 8.0, 10mM MgCl₂, 100mM NaCl, then
 phenol extracted, ethanol precipitated and resuspended in 20mM
 20 Tris-HCl pH 8.0, 12mM MgCl₂, 12mM CaCl₂, 600mM NaCl, 1mM EDTA
 and reacted at 15°C with the exonuclease Bal31. After removal
 of approx 1.8Kb the reaction was stopped by means of the
 addition of EGTA, phenol extracted, ethanol precipitated and
 vacuum dried. The pellet was resuspended in TE, blunt ends
 25 formed with DNA polymerase klenow fragment and ligated in the
 presence of BglII linker.

Digestion with BamHI and HindIII facilitates the insertion of
 either an E. coli trp or tac promoter. The BamHI BglII fragment can
 be used as a cassette for the expression of 7-keto-8-aminopelargonic
 30 acid synthetase in a range of microorganisms including yeasts,
 bacillus, lactobacillus etc.

The DNA sequence of the bio F gene is listed in table 6.

Ligation of the BamHI BglII fragment into pMA91 or pMA36C provided novel plasmids suitable for transformation of yeast. The procedure is shown in Fig 8.

7. Nucleotide and Amino acid sequence of Bio B gene.

5 Confirmation that pB₁OBla contained a functional biotin synthetase gene was by complementation of E.coli JM83-14a, a bioB mutant constructed in our laboratory, grown on a minimal-ampicillin medium plus dethiobiotin. The complete nucleotide sequence and deduced amino-acid sequence are
10 displayed in Table 1.

The Bio B gene comprises 346 codons with a calculated protein molecular weight of 38, 637 daltons. The M_r value is approximately 2,637 larger than that predicted by Dottin et al. (1975) using S.D.S polyacrylamide gel electro-
15 phoresis, a difference probably due to errors involved in M_r determination by this method owing to protein conformation. From our data the Robson conformation for secondary structure of the protein (Garnier et al., 1978) gives helix 45%, extended 36%, turn 9% and coil 10% whilst the Chou-Fassman
20 method (Chou and Fassman, 1978), with its inherent overlap (Nishikawa, 1983), the proportions are helix 57%, extended 72% and turn 29%. These regions are distributed more or less evenly throughout the molecule in tracts of no more than 24 amino-acid residues.

25 Nath and Guha (1982) suggested that translation of the bioB gene could not initiate at the ATG already defined (Otsuka and Abelson, 1978; Barker et al., 1981) because of an in-frame terminating codon, TGA, situated 73 bases downstream. This observation was based on dissertation material (Otsuka,
30 1979) which we have not seen. However, our sequence indicates that although a TGA triplet does exist at approximately that position, it is in fact out of frame.

Similarly, although not directly part of this work, we have also sequenced the regulatory region of the biotin operon (data not shown) and find TGGAGAAGCCCC immediately preceeding the initiating ATG of the bioB gene. This sequence is in agreement with that of Barker et al., (1981) but not with that of Otsuka and Abelson (1978) in which the GA pair (underlined) is missing. Therefore we reaffirm the location of the start of the bioB gene.

Whilst attempts to purify enzymes of the biotin operon from E.coli have found only small amounts of protein (Krell and Eisenberg, 1970; Stoner and Eisenberg, 1975), suggesting that the operon is weakly expressed, which is to be expected because of the very small quantities of biotin ($\sim 10\text{ng L}^{-1}$) required for normal growth of this bacterium, analysis of the codon usage does not reflect this. Although within the bioB coding region there is indeed an 81% base match for weakly expressed E.coli genes, there is also a 78% base match for strongly expressed genes (DNASTAR). If the same analysis is performed for yeast genes, then only a 73% base match is found.

8. Incorporation of E. coli Bio A, B, C, D and F genes into other Yeast expression vectors.

Using techniques analogous to those described above the Bio A, B, C, D and F genes were incorporated into the Yeast expression vector plasmid pKV49 as shown in Fig 10. The BglII expression site shown in pKV 49 is a convenient region for insertion of the indicated gene cassettes containing the particular gene.

Use of Polymerase Chain Reaction to construct cassettes
containing the Biotin Operon genes.

The recently developed Polymerase Chain Reaction (PCR) provides a method for increasing the number of copies of a given DNA sequence, without the need to culture the organism which contains it. The PCR proceeds in three phases.

1. The DNA is denatured.
2. The short oligomer primers are annealed to the DNA flanking the target DNA sequence.
3. The target DNA is enzymatically extended from the primers across the target region using Taq DNA polymerase.

The process is repeated and results in the exponential accumulation of the specific target sequence, approximately 2^n , where n is the number of cycles of melting and primer extension. The method is of use in the construction of specific sequences where one wishes to add particular restriction sites, or modify the ends of the target genes. This is achieved by constructing a primer oligomer, part of which binds to the target gene and which acts as a primer, and the rest of which forms the restriction sites of the adaptor. Specific constructions for each of the Biotin genes are detailed below.

Approximately 1.0 μ g of the plasmid pH7 DNA in water is mixed with the appropriate primers, buffer water 20% Triton and Taq DNA polymerase at the concentrations shown in methods 1, 2 and A - F. The reaction tubes are set up in the PCR reaction programmable heating cooling block at the segment temperatures and times detailed in method 1.

At the end of the reaction cycles the DNA is run on a 1% agarose gel made up in Tris-HCl 0.889 M, boric acid 0.089 M, EDTA 0.002 M, and electrophoresed for approximately 15 hours at 25V. The gel was stained with ethidium bromide (1 μ g/ml) and visualised with UV. The prominent band of DNA at the expected molecular weight is electrophoresed onto DE 81 paper. The DNA is removed from this paper in a high salt buffer comprising NaCl 2M, Tris-HCl pH 8.0 50 mM, and EDTA 1.0 mM. The DNA

is ethanol precipitated using 0.1 volume 3M sodium acetate pH 7.5. 2 volumes of 95% ethanol. left at -20°C for 30 minutes and centrifuged at 12,000g for 30 minutes at 4°C.

The pellet is then washed three times with 70% ethanol and vacuum dried. The DNA is then resuspended to 0.1 ug/ml in TE (10mM Tris HCl, 1mM EDTA pH 8.0). The DNA obtained is cut with the appropriate restriction enzymes and inserted into the plasmid of choice using standard ligation protocols.

Thus the genes may be made suitable for expression in a variety of species. eg yeasts lactobacillus and bacillus. using the appropriate vectors.

The individual gene products. following insertion of the gene alone or in combination, are thus made available to either supplement the enzyme activity already present or to provide an activity which was previously lacking in that particular species.

It should be noted that the start codon of the bio B gene as it occurs in E. coli has been changed from GTG which codes for formyl methionine in E. coli. to ATG which codes for methionine, and now allows expression of this gene in other species. particularly eukaryotes.

Method 1

Programme 1:

		<u>Temp °C</u>	<u>Time mins.</u>
	Segment 1	90	2.0
25	Segment 2	55	2.5
	Segment 3	70	3.0

Polymerase Chain Reaction X 4 Buffer:

	<u>Component</u>	<u>Final Conc.</u>	<u>Stock</u>	<u>Amount / 1ml</u> <u>X 4 mix</u>
30	Tris HCl pH8.3	10.0 mM	1.0 M	40.0 µL
	KCl	50.0 mM	1.0 M	200.0 µL
	MgCl ₂	1.5 mM	1.0 M	6.0 µL
	Gelatin	0.01 %	1.0 %	40.0 µL
	dATP	200 µM	100 mM	2.0 µL
35	dTTP	200 µM	100 mM	2.0 µL

(contd.)

	<u>Component</u>	<u>Final concn.</u>	<u>Stock</u>	<u>Amount / 1ml X 4 mix</u>
5	CTP	200 μ M	100 mM	9.0 μ L
	dGTP	50 μ M	100 mM	2.0 μ L
	dc ⁷ GTP	150 μ M	10.0mM	15.0 μ L

Make up to 1000 μ L with sterile distilled water. use 25 μ L per 100 μ L reaction.

- 10 Oligonucleotide Synthesis:
Oligonucleotides were synthesised using cyanoethyl phosphoramidite method on an Applied Biosystems 381A synthesiser at 0.2 μ M scale.

Method A

- 15 Bio A gene from E coli: construction of cassette for expression in heterologous organism.

oligo 1 = 5' gcggccgcgaattcagatctataatgacaacggacgatcttgcc 3'

oligo 2 = 5' gcggccgcaagcttagatctttattggcaaaaaaatgtttca 3'

Bio A gene oligo 1 44/mer 2253 μ g/ml

- 20 oligo 2 42/mer 2236 μ g/ml

want final concentration of approximately 20 pmol dilute:

oligo 1 1:15 66 μ L of oligo + 933.4 μ L of water = 10 pmol = 145ng/ μ L

oligo 2 1:21 47.6 μ L of oligo + 952.4 μ L of water = 10 pmol = 133ng/ μ L

PCP

- 25 X4 Buffer 25 μ L
primer 1 2.0 μ L
primer 2 2.0 μ L
10% triton 1.0 μ L
30 DNA 10.0 μ L (approx 10 μ g)
Taq 1.0 μ L (5.0 U)
Water 59 μ L
Total 100 μ L
Overlay 100 μ L sterile mineral oil
25 25 x cycles
programme 1

(DNA = biotin operon Eco/Eco fragment of 1H7)

Method B

Bio B gene from E. coli: construction of cassette for expression in heterologous organism.

oligo 1 = 5' gcggccgcgaattcagatctataatggctcaccgcccacgctgg 3'

5 oligo 2 = 5' gcggccgcaagcttggatcctcataatgctgccgcgttgtaa 3'

Bio B gene oligo 1 44/mer 2153 µg/ml

oligo 2 42/mer 2279 µg/ml

want final concentration of approximately 20 pmol

dilute:

10 oligo 1 1:15 66.6 µl of oligo + 933.4 µl of water = 10 pmol = 145 ng/µl

oligo 2 1:15 62.5 µl of oligo + 937.5 µl of water = 10 pmol = 138 ng/µl

PCR

	X 4 Buffer	25 µl
	primer 1	2.0 µl
15	primer 2	2.0 µl
	10 % triton	1.0 µl
	DNA	10.0 µl (approx 1.0 µg)
	Taq	1.0 µl (5.0 U)
	Water	59 µl
20	Total	100 µl
	overlay 100 µl of sterile mineral oil	
	25x cycles	
	programme 1	

(DNA = biotin operon Eco/Eco fragment of 1H7)

25

Method C

Bio C gene from E. coli: construction of cassette for expression in heterologous organism.

oligo 1 = 5' gcggccgcgaattcggatccataatggcaacgggtaataaaca 3'

oligo 2 = 5' gcggccgcaagcttggatccttactcacgagcaatcactcc 3'

30 Bio C oligo 1 44 mer 2617 µg/ml

oligo 2 41 mer 2163 µg/ml

want final concentration of approximately 20 pmol

dilute:

oligo 1 1:18 55.5 µl of oligo + 944.5 µl of water = 10 pmol = 145 ng/µl

35 oligo 2 1:16 62.5 µl of oligo + 937.5 µl of water = 10 pmol = 135 ng/µl

PCR

5 X4 buffer 25 µl
 primer 1 2.0 µl
 primer 2 2.0 µl
 10% triton 1.0 µl
 DNA 10.0 µl (approx 1.0 µg)
 Taq polymerase 1.0 µl (5.0 U)
 Water 59 µl
 Total 100 µl
 overlay 100 µl of sterile mineral oil
 10 25 x cycles
 programme 1

(DNA = biotin operon Eco/Eco fragment of 1H7)

Method D

15 Bio D gene from E. coli: construction of cassettes for
 expression in heterologous organisms

oligo 1 = 5' tctagaattcggatccataatgagtaaacgttattttgtca 3'

oligo 2 = 5' tctagaagcttagatctacaacaaggcaaggttatgt 3'

Bio D oligo 1 (left) 41 mer 2391 µg/ml

oligo 2 (right) 38 mer 2292 µg/ml

20 want final concentration of approximately 20 pmol
 dilute:

oligo 1 1.15 µl of oligo = 20 pmol = 274 ng/µl

oligo 2 1.1 µl of oligo = 20 pmol = 254 ng/µl

PCR

25 X4 buffer 25 µl
 primer 1 1.15 µl
 primer 2 1.1 µl
 10% triton 1.0 µl
 DNA 10.0 µl (approx 5.0 µg)
 30 Taq polymerase 1.0 µl (5.0U)
 water 60.75 µl
 total 100 µl
 overlay 100 µl of sterile mineral oil
 25 x cycles
 35 programme 1

(DNA = biotin operon Eco/Eco fragment of 1H7)

Method F

Bio F gene from E. coli: construction of cassette
for expression in heterologous organism.

oligo 1 = 5' aagcttggatccataatgagctggcaggagaaaatcaacgoggc 3'
5 oligo 2 = 5' cagctgcagatctttaaccgttgccatgcagcacctccagca 3'

Bio F gene oligo 1 44/mer 2040 µg/ml

oligo 2 42/mer 2000 µg/ml

want final concentration of approximately 20 pmol

dilute:

10 oligo 1 1:14 71.5µl of oligo + 928.5 µl of water = 10 pmol = 145ng/µl

oligo 2 1:22 45.5µl of oligo + 954.5 µl of water = 10 pmol = 139ng/µl

PCR

	X ⁴ buffer	25 µl
	primer 1	2.0 µl
15	primer 2	2.0 µl
	10% triton	1.0 µl
	DNA	10.0 µl (approx. 1.0µg)
	Taq polymerase	1.0 µl (5.0 U)
	Water	59 µl
20	Total	100 µl
	overlay 100 µl of sterile mineral oil	
	25 x cycles	
	programme 1	

(DNA = biotin operon Eco/Eco fragment of LH7)

25

30

References

- Sanger, F, Nicklen, S and Coulson A R. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463-5467.
- Tautz, D and Renz, M: Optimised Freeze-squeeze method for recovery of DNA fragments from agarose gels. Anal Biochem 132 (1983) 14-19.
- Wright, L D. and Skeggs, E R: Determination of biotin with Lactobacillus arabinosus. Proc. Soc. Exp. Med 56 (1944) 95-102.
- Dottin, R P, Cutler, L S and Pearson M L: Repression and autogenous stimulation in vitro by bacteriophage lambda repressor. Proc. Nat. Acad. Sci. USA 72 (1975) 084-808.
- Garnier, J, Osugthorpe, D and Robson, B: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120 (1978) 97-120.
- Chou, P P, and Fassman G D: Prediction of the secondary structure of proteins from their amino-acid sequence. Adv. Enzymol. 46 (1978) 45-148.
- Nishikawa, K: Assessment of secondary-structure prediction of proteins, comparison of computerized Chou-Fassman method with others. Biochim. Biophys. Acta 748 (1983) 285-199.
- Nath, S K and Guha A: Abortive termination of bioBFCD RNA synthesized in vitro from the bio ABFCD operon of Escherichia coli K-12. Proc. Natl. Aced. Sci. USA 79 (1982) 1786-1790.
- Otsuka, A and Abelson J: The regulatory region of the biotin operon in Escherichia coli. Nature 276 (1978) 689-693.
- Barker, D F. Kuhn J and Campbell, A M: Sequence and properties of operator mutations in the bio operon of Escherichia coli. Gene 13 (1981) 89-102.
- Otsuka, A: Ph.D Thesis, University of California, San Diego, 1979.
- Krell, K and Eisenberg, M A: The purification and properties of dethiobiotin synethase. J. Biol. Chem 245 (1970) 6558-6566.

Stoner, G L and Eisenberg, M A: Purification and properties of 7,8,-diaminopelargonic acid aminotransferase. J. Biol. Chem. 250 (1975) 4029-4036.

Cohen, G, Zimmer A, Gurevitch E, and Yankofsky S, : Isolation and Characterisation of a ColE1 plasmid containing the entire bio gene cluster of Escherichia coli K-12. Mol. Gen. Genet. 166 (1978).

Cleary P P, Campbell A, and Chang E: Location of promoter and operator sites in the biotin gene cluster. Proc. Natl. Acad. Sci. USA 69 (1972) 2219-2223.

Guha A, Saturen Y, and Szybalski W : Divergent orientation of transcription from the biotin locus of Escherichia coli. J. Mol. Biol. 56 (1971) 53-62.

Eisenberg M A : Biotin: biogenesis, transport and their regulation. Adv. Enzymol. 33 (1973) 317-372.

Niimura T, Suzuki R and Sahashi Y : Studies on the formation of biotin from dethiobiotin and sulphate in Saccharomyces cerevisiae J. Vitaminol. 10 (1964) 218-223.

Das Gupta C K, Guha A (1978) Isolation of the regulatory segment of the Biotin operon of E. coli K-12. Gene 3. p232-246.

Viera J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers.

Bachmann B S, Low F B (1976) Linkage map of E. coli K-12, edition 6 Microbiol. Rev., 40, 1-56.

Table 1

E.coli Bio B gene with sequence modified
for Yeast.

1	ATGGCTCATAGACCAAGATGGACTTTGTCTCAAGTTACTGAATTGTTCTGA	50
51	AAAGCCATTGTTGGACTTGTGTTCTGAAGCTCAACAAGTTTCACAGACAAC	100
101	ACTTCGACCCAAGACAAGTTCAAGTTTCTACTTTGTTGTCTATCAAACT	150
151	GGTGCTTGTCCAGAAGACTGTAAGTACTGTCCACAATCTTCTAGATACAA	200
201	GACTGGTTTGGAAAGCTGAAAGATTGATGGAAGTCGAACAAGTTTTGGAAT	250
251	CTGCTAGAAAGGCTAAGGCTGCTGGTTCTACTAGATTCTGTATGGGTGCT	300
301	GCTTGGAAAAACCCACACGAAAGAGACATGCCATACTTGGAACAAATGGT	350
351	TCAAGGTGTTAAGGCTATGGGTTTGGAAAGCTTGTATGACTTTGGGTACTT	400
401	TGTCCGAATCCCAAGCTCAAAGATTGGCTAACGCTGGTTTGGACTACTAC	450
451	AACCACAACCTTGGACACTTCTCCAGAATTCTACGGTAACATTATTACTAC	500
501	TAGAACCTACCAAGAAAGATTGACACCTTGGAAAAGGTTAGAGACGCTG	550
551	GTATTAGGTTTTGTTCCGGTGGTATCGTTGGTTTGGGTGAAACTGTTAAG	600
601	GACAGAGCTGGTTTTGTTGTTGCAATTGGCCAACTTGCCAACCCACCAGA	650
651	ATCTGTTCCAATTAACATGTTGGTTAAGGTTAAGGGTACTCCATTGGCTG	700
701	ACAACGACGACGTTGACGCTTTCGACTTCATTAGAACTATCGCTGTTGCT	750
751	AGAATTATGATGCCAACTTCCTACGTTAGATTGTCCGCTGGTAGAGAAACA	800
801	AATGAACGAACAAACCCCAAGCTATGTGTTTCATGGCTGGTGCTAACTCTA	850
851	TCTTCTACGGTTGTAASTTGTGACTACTCCAAACCCAGAAAGACAAAG	900
901	GATTTGCAATTGTTGAGAAAGTTGGGTTTGAACCCACAACAACCGCTGT	950
951	CTTGGCTGGTGACAACGAACAACAACAAGATTGGAACAGGCTTTGATGA	1000
1001	CTCCAGATACTGATGAATACTACAACGCTGCTGCTTTGTAG	1041

Table 2

THE 5104 GENE OF E. coli: DNA SEQUENCE, DERIVED AMINO ACID SEQUENCE AND
PRIMERS USED TO CONSTRUCT THE GENE CASSETTE.

5' : 3' :

```

gggggggggagtttcagatctatcATGACGACGGACGATCTTGGCTTTGACCAACCGATATCTGGCAGCCATACACATCCATG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
MetThrThrAspAspLeuAlaPheAspGlnProHisIleTrpHisProTyrThrSerMet
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACCTCCGCTGTGGCGCTTTATCCGGTGGTGGCGCGCGAAGGTTGGCGAGTGGATTTTGTCT
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ThrSerProLeuProValTyrProValValSerAlaGluGlyCysGluLeuIleLeuSer
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GACGGCAGACGCTTGGTTGACGGTATGTCGTCTGGTGGCGGGCGATCCACGGCTACAAAT
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
AspGlyArgArgLeuValAspGlyMetSerSerTrpTrpAlaAlaIleHisGlyTyrAsn
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CAGCGCTAGCTTAATCGCGCGATGAGTGGCAATTCATGCGATGTCGGATGTGATGTTT
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
HisProGlnLeuAsnAlaAlaMetLysSerGlnIleAspAlaMetSerHisValMetPhe
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GTCGTATCAGCCATGGCGACCGCATTTGAGGTGTGCGCAACTGGTGGCGATGAGCGCG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GlyGlnIleThrHisAlaProAlaIleGluLeuCysArgLysLeuValAlaMetThrPro
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CGACCGCTGGAGTGGCTTTTCTCGCGCATGGGTTCCGTAACGGTGGAGTGGCGATG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GlnProLeuGluCysValPheLeuAlaAspSerGlySerValAlaValGluValAlaMet
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
AAAAATGGCTTGGAGTACTGGCAAGCCAAAGCGAAGCGCGCAGCGCTTTTCTGACITTC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
LysMetAlaLeuGlnTyrTrpGlnAlaLysGlyGluAlaArgGlnArgPheLeuThrPhe
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CGCAATGGTTATCATGGCGATACCTTTGGCGCGATGTGGGTGTGGCATCCGGATAACTCA
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ArgAsnGlyTyrHisGlyAspThrPheGlyAlaMetSerValCysAspProAspAsnSer
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ATGCACAGTGTGTGGAAAGGCTACCTGGCAGAAACCTGTTTGGTCCCGCCCGCGAAAGC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
MetHisSerLeuTrpLysGlyTyrLeuProGluAsnLeuPheAlaProAlaProGlnSer
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CGCATGGATGGCGAATGGGATGAGCGCGATATGGTGGGCTTTGGCGCGCTGATGGCGCG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ArgMetAspGlyGluTrpAspGluArgAspMetValGlyPheAlaArgLeuMetAlaAla
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GATCGTCATGAATCGCGCGGTGATCATTGAGCGCATTTGCCAGGGCGCAGCGCGGATG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
HisArgHisGluIleAlaAlaValIleIleGluProIleValGlnGluAlaGlyGlyMet

```

Table 2 (contd.)

```

CGATGTACCATCGGGAATGTTAAACGAAATCGGCAAAATATCGGATCGGCAAGGATG
-----
ArgMetIleHisProGluTyrLeuLeuArgIleArgGluIleCysAspArgGluGlyIle
-----
TTGCTGATTGCGAGCGAGATCGGCACTGCGATTGCTGCTACCGGGAATGCTTTGCTGT
-----
LeuLeuIleAlaAspGluIleAlaThrGlyPheGlyArgThrGluValLeuPheAlaCys
-----
GAACATGCAAGAAATCGGCGGCGACATTTTGTGCTGCTGCTAAAGCGTTAAGCGGCGCA
-----
GluHisAlaGluIleAlaProAspIleLeuGlyLeuGlyValAlaLeuThrGlyGlyThr
-----
ATGACCGCTTTTGGCGCACTGACCGCGCGAGGTTGCAAGAAACATGAGTACCGGTGAA
-----
MetThrLeuSerAlaThrLeuThrThrArgGluValAlaGluThrIleSerAsnGlyGlu
-----
GCGCGTTGCTTTATGCGATGCGGCAACTTTTATGCGCAATCGGCTGCGCTGCGCGCA
-----
AlaGlyCysPheMetHisGlyProThrPheMetGlyAsnProLeuAlaCysAlaAlaAla
-----
AAGCGGAGCGCTGCGGATTGCGAATGCGGCACTGCGGCAAGCGGCTGCGGATATTGAA
-----
AsnAlaSerLeuAlaIleLeuGluSerGlyAspTrpGluGluValAlaAspIleGlu
-----
GTACAGCTGCGGCAAGCAACTTGGCGCGCGGCTGATGCGGAAATGCTTGGCGATGTGCG
-----
ValGluLeuArgGluGluLeuAlaProAlaArgAspAlaGluMetValAlaAspValArg
-----
GTACTGCGGCGGCAATGCTGCTGCTGAAACCACTCATCGGCTGAAATATGCGGCGCTGCA
-----
ValLeuGlyAlaIleGluValValGluThrThrHisProValAsnMetAlaAlaLeuGlu
-----
AAATGCTTTGCGAAGCGGCTGCTGCAATCGGCGGCTTTTGGCAAGCTGATTTACCTGATG
-----
LysPhePheValGluGluGlyValTrpIleArgProPheGlyLysLeuIleTyrLeuMet
-----
CGGCGCTATATTATCTCGCGCAAGCACTTGCAGCGCTGCGCGCGCGGTTAAGCGCGCG
-----
ProProTyrIleIleLeuProGluGluLeuGluArgLeuThrAlaAlaValAsnArgAla
-----
GTACAGGATGAACATTTTTTGCATTAAGatcttaagcttgcgccgc
-----
ValGluAspGluThrPhePheCysGluEnd  BglII HindIII

```

```

5' start 3'
Clp01 44mer at start GCGGCGCGCAATTCAGATCTATATGACACCGGACGATCTTGGC
-----
EcoRI BglII

```

```

5' stop 3'
Clp02 42mer at end GCGGCGCGCAAGCTTAGATCTTTATGCGCAAGAAATGTTTCA
-----
HindIII BglII

```

Table 2

THE *hcd* GENE OF *E. coli*: DNA SEQUENCE, DERIVED AMINO ACID SEQUENCE AND PRIMER USED TO CONSTRUCT THE GENE CASSETTE.

Ec:PI SpIII

```

ggggctggcagctcagatcttatatggctcaccgcccagcgtggacatttctggcaagtcacagattatttgaaaaacgcttg
-----
MetAlaHisArgProArgTrpThrLeuSerGlnValThrGluLeuPheGluLysProLeu
-----
ctggatctctctcttttgaagcgacagcagctgcatcgccagcattttcgaatctctctcagctt
-----
LeuAspLeuLeuPheGluAlaGlnGlnValHisArgGlnHisPheAspProArgGlnVal
-----
cagctcagcagcttgcctgtcgtatgaagaccgagcttgtccggaagattgcaatactgc
-----
GlnValSerThrLeuLeuSerIleLysThrGlyAlaCysProGluAspCysLysTyrCys
-----
ccgcfaagctcgctacaaacccggctggaagccgagcggcttcttggaaattgaaacg
-----
ProGlnSerSerArgTyrLysThrGlyLeuGluAlaGluArgLeuMetGluValGluGln
-----
gtcttgagctggcgccgcaaacgcaagcgcagcagctcagcgcctttctgtatgagcgcg
-----
ValLeuGluSerAlaArgLysAlaLysAlaAlaGlySerThrArgPheCysMetGlyAla
-----
ggctgcaaaatccccacgaacgcatatgcccatactggaacaaatgctgcagggggtt
-----
AlaTrpLysAsnProHisGluArgAspMetProTyrLeuGluGlnMetValGlnGlyVal
-----
aaagtcagtgggctggagcgctgtatgacgctgggcagcttgagtgaaatctcagggcag
-----
LysAlaMetGlyLeuGluAlaCysMetThrLeuGlyThrLeuSerGluSerGlnAlaGln
-----
cgcctgcccgaagcccgccctggatttactacaacacaaacctgacacatctggccggatct
-----
ArgLeuAlaAsnAlaGlyLeuAspTyrTyrAsnHisAsnLeuAspThrSerProGluPhe
-----
tacggcaatattcattcaccacacgcaattatcaggaacccctgatatgctggaaaaagt
-----
TyrGlyAsnIleIleThrThrArgThrTyrGlnGluArgLeuAspThrLeuGluLysVal
-----
cgcgatcgccgatcaaatgtctgtctgscggcatttctgggcttagggcaaacggtaaaa
-----
ArgAspAlaGlyIleLysValCysSerGlyGlyIleValGlyLeuGlyGluThrValLys
-----
gatacgcccgattattgctgcaactggcaaacctgcccagccgcccggaaagcgtgccc
-----
AspArgAlaGlyLeuLeuLeuGlnLeuAlaAsnLeuProThrProProGluSerValPro
-----
atcaacatctctgtgaagctgcaaacgacgccccttgcgataagcagatgctgcatgct
-----
IleAsnMetLeuValIleValLysGlyThrProLeuAlaAspAsnAspAspValAspAla

```

Table 3 (contd.)

```

TTTGTATTTATTCGCPCCATTGCGGTCGCGCGGATCATGATGCCAAGCTCTTACGTCGCT
-----
PheAspPheIleArgThrIleAlaValAlaArgIleMetMetProThrSerTyrValArg
-----
CTTTCTGCGGACGCGGAGCATGACGACACAGACTCAGGCGATGTCCTTTATGCGAGGCT
-----
LeuSerAlaGlyArgGluGlnMetAsnGluGlnThrGlnAlaMetCysPheMetAlaGly
-----
GCAAACTCGATTTTCTACGCTTSCAAACTGCTGACACGCGGATGCGGAGGAGATAA
-----
AlaAsnSerIlePheTyrGlyCysLysLeuLeuThrThrProAsnProGluGluAspLys
-----
GACCTGCAACTGTTCCGCAAACTGSSGCTAAATCCGAGCAAACTGCGCTGCTGCGAGG
-----
AspLeuGlnLeuPheArgLysLeuGlyLeuAsnProGlnGlnThrAlaValLeuAlaGly
-----
GATAACGAAACAACGCAACGCTCTTGAACAGGCGCTGATGACCCGAGACCGACGCAATAT
-----
AspAsnGluGlnGlnGlnArgLeuGluGlnAlaLeuMetThrProAspThrAspGluTyr
-----
TACAACGCGGCAEDATTATGAggatcgaagcttgcggccgc
-----
TyrAsnAlaAlaAlaLeuEnd  BamHI  HindIII

```

```

              EcoRI  BglII
5' / / start 3'
ClpEI 48bp at start  GCGGCGCGCAATTCAGATCTATAATGCGCTCACCAGCCACGCTGG
              stop 3'
5' / / stop 3'
ClpEI 48bp at end  GCGGCGCGCAAGCTTGGATCCGATAATGCTGCGGCGCTTGTAA
              HindIII  BamHI

```

5561 5562

Met-Ala-Thr-Val-Asn-Leu-Gln-Ala-Ile-Ala-Ala-Ala-Phe-Gly-Arg-Ala-Ala-Ala-His-Tyr
 GAGGACAGATGCAGATCTACAGCGGCCAGAGTGTGACGGCCTTACTGGCAATGCTTCACACAG
 Glu-Gln-His-Ala-Asp-Leu-Gln-Arg-Gln-Ser-Ala-Asp-Ala-Leu-Leu-Ala-Met-Leu-Pro-Gln
 CATAAATTACAGCCACGTAAGTGTGACGGCGGGTTTGTSSACCTGGCTGGATGAAGCCGGCACTGG
 Arg-Lys-Tyr-Trp-His-Val-Leu-Asp-Ala-Gly-Cys-Gly-Pro-Gly-Trp-Met-Ser-Arg-His-Trp
 CGSSAACGCTACGGCGCAGTGACAGCGCCTTAGATCTCTGGCCGCCAAATGCTTGTTCAGGCA
 Arg-Glu-Arg-His-Ala-Gln-Val-Thr-Ala-Leu-Asp-Leu-Ser-Pro-Pro-Met-Leu-Val-Gln-Ala
 CCTCAGACGATGCCCGACACDATTATCTGGTGGAGCATATCGAATCCCTGGCCTTAGCG
 Arg-Gln-Lys-Phe-Ala-Ala-Asp-His-Tyr-Leu-Ala-Gly-Asp-Ile-Glu-Ser-Leu-Pro-Leu-Ala
 ACTGCGACGCTTGGATCTTSCATGSAACAATCTCGAGTGCAGTGGTGGAGTAATTTATCC
 Thr-Ala-Thr-Phe-Asp-Leu-Ala-Trp-Ser-Asn-Leu-Ala-Val-Gln-Trp-Cys-Gly-Asn-Leu-Ser
 AGGGCACTCGCGGAGCTGTATCGGTGGTGGCGCCCCAAGGGCTGGTGGCGTTTTACCGAG
 Thr-Phe-Leu-Arg-Glu-Leu-Tyr-Arg-Val-Val-Arg-Pro-Lys-Gly-Val-Val-Ala-Phe-Thr-Thr
 CTGGTGGACGGATGTTACCGCACTGCAACAGGCTGGCAGGCGGTTGACGAGAGGCTCG
 Leu-Val-Gln-Gly-Ser-Leu-Pro-Glu-Leu-His-Gln-Ala-Trp-Gln-Ala-Val-Asp-Glu-Arg-Pro
 CATGCTAATGCTTTTACCGGCCAGATGAAATCGAACAATGCTGTAACGGCGTGCATTAT
 His-Ala-Asn-Arg-Phe-Leu-Pro-Pro-Asp-Glu-Ile-Glu-Gln-Ser-Leu-Asn-Gly-Val-His-Tyr
 CAGCATCATATTGASGCCATCACGCTGTGGTTTGTATGATGCGCTCACTGGCATGCGTTCS
 Gln-His-His-Ile-Phe-Pro-Ile-Thr-Leu-Trp-Phe-Asp-Asp-Ala-Leu-Ser-Ala-Met-Arg-Ser

Table 4 (contd.)

```

CTGAACAGGATCGCTTCCAGCCATCTTCATGAAGGCGCGACCTGCGAATATTAACTGCT
-----
LeuLysGlnIleGlyAlaThrHisLeuHisGluGlyArgPspPrrArgIleLeuTrpArg
TGGCACTTGCAGCGATTGCACCTGGCTTGGCCGCAACACGACGCGGCGATATCTCTGACG
-----
SerGlnLeuGlnArgLeuGlnLeuAlaTrpPrrGlnGlnGlnGlyArgTyrProLeuThr
TATCATCTTTTTTTTGGCACTGATTGCTCGTGAAGTARAggatccaaagcttgcggcgcc
-----
TyrHisLeuPheLeuGlyValIleAlaArgGluEnd  BamHI  HindIII

```

```

              EcoRI  BamHI
5'      /      /      start      3'
Oligo1 primer at start  GCGGCGCGCAATTCGGATCCCTAATGCCAAGCGTTAATAACAA
              HindIII  BamHI
5'      /      /      stop      3'
Oligo2 primer at start  GCGGCGCGCAAGCTTGGATCCCTTACTCAGAGCAATCACTCC

```

Table 5

THE *PIG* GENE OF *E. coli*: DNA SEQUENCE, DERIVED AMINO ACID SEQUENCE AND PRIMERS USED TO CONSTRUCT THE GENE CASSETTE.

StoRI BamHI

```

tctagacttcggatccctaatTGAATAACGTTATTTTGTACCGAACGCGATACCGAAGTGGGGAAAACTGTCGCCAGT
-----
MetSerLysArgTyrPheValThrGlyThrAspThrGluValGlyLysThrValAlaSer
-----
TGTGCATTTTACAGGCGCAAGGCGACGAGGCTACCGGACGGCAGGTTATAAAGCGGTC
-----
CysAlaLeuLeuGlnAlaAlaLysAlaAlaGlyTyrArgThrAlaGlyTyrLysProVal
-----
GCCTCTGGCAGCGAAGAGACCCCGGAAGGTTTACGCAATAGCGACGCGCTGGCGTTACAG
-----
AlaSerGlySerGluLysThrProGluGlyLeuArgAsnSerAspAlaLeuAlaLeuGln
-----
CGCAACAGCAGGCTGCAAGCTGGATTACGCAACAGTAATGCTTACAGCTTCGCAAGACCC
-----
ArgAsnSerSerLeuGlnLeuAspTyrAlaThrValAsnProTyrThrPheAlaGluPro
-----
AGTTGGCGGACATCTCAGCGCGCAAGAGGGCAGACCGATAGATCATTGGTAATGAGC
-----
ThrSerProHisIleIleSerAlaGlnGluGlyArgProIleGluSerLeuValMetSer
-----
GCGGATTACGCGCGCTTGAACACAGGCTGACTGGGTGTTAGTGGAAAGGTGCTGGCGGT
-----
AlaGlyLeuArgAlaLeuGluHisLysAlaAspTrpValLeuValGluGlyAlaGlyGly
-----
TGGTTTACGCGGCTTTCTGACATTTTCACTTTTGCAGATTGGGTAAACACAGGAACAATG
-----
TrpPheThrProLeuSerAspThrPheThrPheAlaAspTrpValThrGlnGluGlnLeu
-----
CGGTTGATAGTGGTGGTGGTGAAGCTGGGCTGTATTAAATCAGCGCATGTTGACTGCA
-----
ProValIleLeuValValGlyValLysLeuGlyCysIleAsnHisAlaMetLeuThrAla
-----
CAGGTAAATCAACACGCGGACTGACTCTGCGCGGTTGGGTGGCGAAGCATGTTACGCGT
-----
GlnValIleGlnHisAlaGlyLeuThrLeuAlaGlyTrpValAlaAsnAspValThrPro

```

Table 5(cont'd.)

```

CTGGGAAACGCTCAGGCTGAATATATGACGAGGCTGACGCGGATGATTCGGCTGGCTG
-----+-----
ProGlyLysArgGluAlaGluTyrMetThr-ThrLeuThrArgMetIleProAlaProLeu
CTGGGAGAGATTCGGCTGGCTGCGAAGATCCAGAAATGCGCGCAACCGGAAAGTGCATA
-----+-----
LeuGlyGluIleProTrpLeuAlaGluAsnProGluAsnAlaAlaThrGlyLysTyrIle

AACTTGGCTTGTGTATGAGTctctaggttctaga
-----+-----
AsnLeuAlaLeuLeuEnd BglII HindIII

```

```

      5'                start                3'
Oligo1 41mer at start TGTAGAATTCGGATCCATAAGAGTAACGTTATTTTGTCA
      / / /
      XbaI EcoRI BsaHI

```

```

      5'                stop                3'
Oligo2 33mer at end TGTAGAAGCTTAGATGTACAGCAAGGCAAGGTTTATGT
      / / /
      XbaI HindIII BglII

```

3.2.1

aggttggattcctatATGAGCTGTGCACGGAGAATTCAACGCCGCGCTCCGATGCGCAGCGTGTGTCGCSATGCCGCTT

 MetSerTrpGlnGluLysIleAsnAlaAlaLeuAspAlaArgArgAlaAlaAspAlaLeu
 CACTGCCGCTTATCGGTTGGCGCACAGGAGCCGACGCTGCGTGCTGCTGCGCATGATCGCCAG

 ArgArgArgTyrProValAlaGlnGlyAlaGlyArgTrpLeuValAlaAspAspArgGln
 TATCTGAACCTTTTCCAGTAACGATTATTTAGGTTTTAAGCCATCATCCGCAAAATTATCCGT

 TyrLeuAsnPheSerSerAsnAspTyrLeuGlyLeuSerHisHisProGlnIleIleArg
 GCTGCGCAGCGAGGGGCGGAGCAATTTGGCATCGGTAGCGGGCGGCTCCGGTCAGCTCAGC

 AlaTrpPlnGlnGlyAlaGluGlnPheGlyIleGlySerGlyGlySerGlyHisValSer
 GCTTATAGCGTGTGTGCATCAGCCACTGSAAGGAGAGGCTGGCGAGTGGCTTGGCTATTCCG

 GlyTyrSerValValHisGlnAlaLeuGluGluGluLeuAlaGluTrpLeuGlyTyrSer
 CGGGCACTGCTGTTTATCTCTGTTTTGCGCGCTAATCAGGCACTTATTGGCGCGATGATG

 ArgAlaLeuLeuPheIleSerGlyPheAlaAlaAsnGlnAlaValIleAlaAlaMetMet
 GCSPAAGAGGAGCGGTATTGCTGCGACCGGCTTAGCCATGGCTCATTGCTGGAAGCTGGC

 AlaLysGluAspArgIleAlaAlaPecArgLeuSerHisAlaSerLeuLeuGluAlaAla
 AGTTTAGGCGCGTCCGAGCTTCCGCGTTTTGCTCATACGATGTCAGTCACTTATTGGCGCGA

 SerLeuSerProSerGlnLeuArgArgPheAlaHisAsnAspValThrHisLeuAlaArg
 TTGCTTGGCTTCCCCCTGTCCGGGGCAGCAAAATGGTGGTGCACAGAAGGCGTETTCAGCATG

 LeuLeuAlaSerProCysProGlyGlnGlnMetValValThrGluGlyValPheSerMet
 GACGGCGATAGTGGCGCACTGGCAATCCAGCAGGTAACGCAACAGCACAATSSCTGGTTG

 AspGlyAspSerAlaProLeuArgIleGlnGlnValThrGlnGlnHisAsnGlyTrpLeu
 ATGGTCGATGATGCCACGGCACGGCGCTTATCGGGCAGCAGGGGCGCGGCGAGCTGCTGG

 MetValAspAspAlaHisGlyThrGlyValIleGlyGluGlnGlyArgGlySerCysTrp
 CTCGCAAGGCTAAGACCASAATTGCTGGTAGTGACTTTTGGCAAGGATTTGGCGTCAGC

 LeuGlnLysValIleProGluLeuLeuValValThrPheGlyLysGlyPheGlyValSer

5' stop 3'
 Oligo6 -2mer at end CAGCTGCAGATCTTTTAACCGTTGCCATSCAGCACCCTCCASCA
 / //
 PvuII / PstI
 /
 BclII